DIETARY PATTERNS IN THE ETIOLOGY OF LYMPHOID NEOPLASMS

Marta Solans Margalef

Universitat de Girona

Doctoral Thesis | 2019

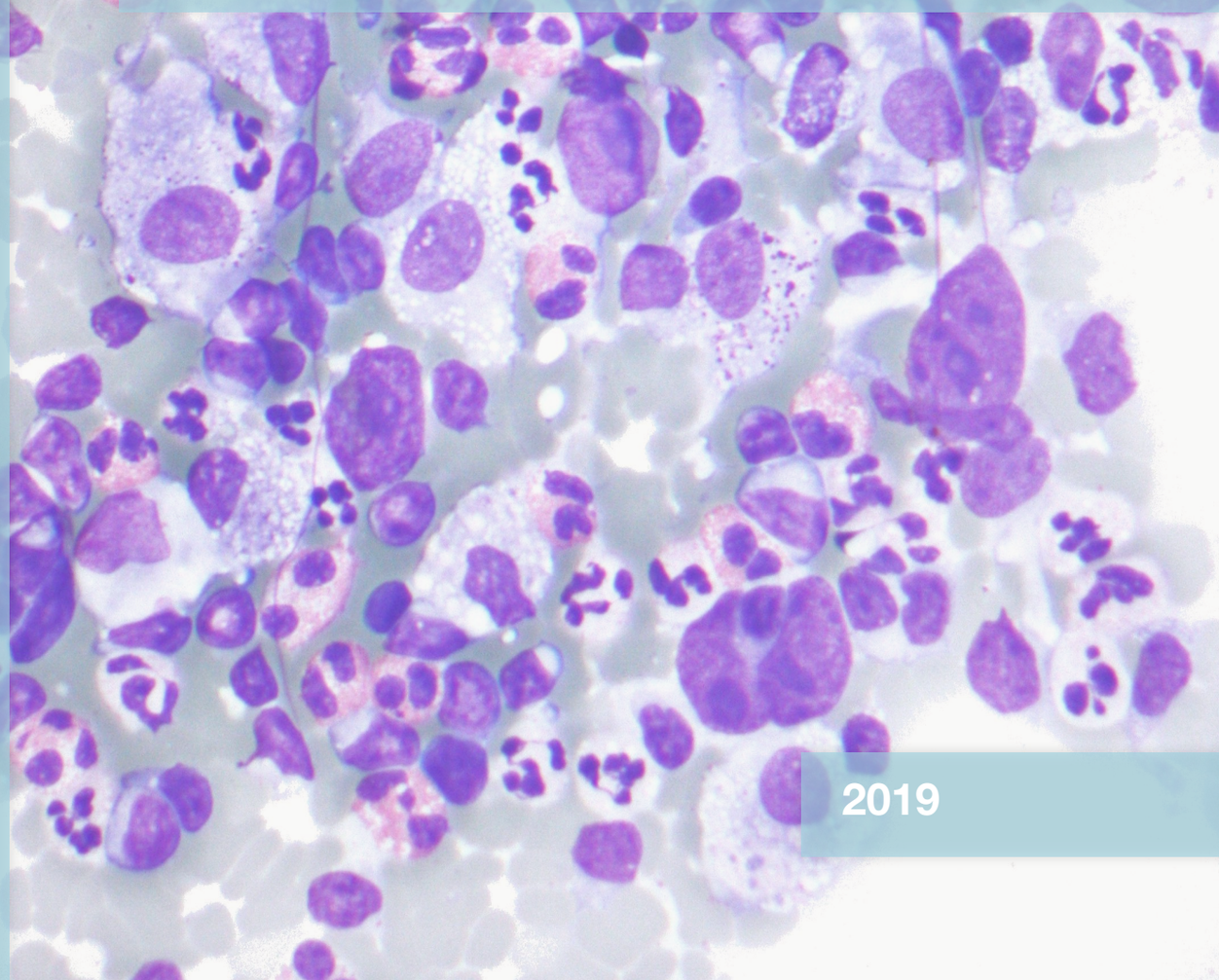
Universitat de Girona

DOCTORAL THESIS

Dietary patterns in the etiology of lymphoid neoplasms

Marta Solans Margalef

Marta Solans Margalef



2019

Universitat de Girona

DOCTORAL THESIS

Dietary patterns in the etiology of lymphoid neoplasms

Marta Solans Margalef

2019

PhD program in Molecular Biology, Biomedicine and Health

Under the direction of:

Prof. Marc Saez Zafra and Dr. Rafael Marcos Gragera

Thesis delivered to obtain the doctoral degree by the Universitat de Girona

El **Prof. Marc Saez Zafra**, de la Universitat de Girona i membre del Grup de Recerca en Estadística, Econometria i Salut (GRECS) i el **Dr. Rafael Marcos Gragera**, de la Universitat de Girona i epidemiòleg de la Unitat d'Epidemiologia i Registre de Càncer de Girona (UERCG) de l'Institut Català d'Oncologia (ICO),

CERTIFIQUEN:

Que el treball titulat "Dietary patterns in the etiology of lymphoid neoplasms", que presenta Marta Solans Margalef per a l'obtenció del títol de doctora, ha estat realitzat sota la seva direcció i que compleix els requisits per poder optar a Menció Internacional.

I, perquè així consti i tingui els efectes oportuns, signem aquest document.

Prof. Marc Saez Zafra

Dr. Rafael Marcos Gragera

Girona, 4 de juliol de 2019

ACKNOWLEDGEMENTS

Gràcies als meus directors de tesi, dels qui tant he après! **Rafa**, no hagués pogut trobar un director més divertit i entregat; gràcies per la teva gran generositat, confiança cega i per haver-me conduït aquell primer dia al despatx d'en Marc. **Marc**, gràcies per la teva immensa dedicació, pel teu suport incondicional i per haver-me cuidat tant.

Gràcies al **GRECS** (Àngels, Anna, Carme, Diego, Elena, Gemma, Germà, Laia, Laura S, Laura V, Lluís i Maria Antònia), la meva família acadèmica! En especial, moltes gràcies **Germà** per haver-me introduït al món del CoDa i de la cuina catalana! Gràcies **Jorge** i **Geòrgia** pel vostre suport i companyia durant tot aquest temps. Ha estat un regal poder compartir aquesta etapa amb gent tan bona però, sobretot, amb tan bona gent!

Gràcies a tots els companys del **Registre de Càncer** (Àngel, Anna, Aina, Carme, Gemma, Joana, Loreto, Montse i Raquel) per haver-me fet sentir com a casa. No m'hagués pogut divertir més anant de "bolos", als vostres pica-piques i a les temudes sessions d'hematologia!

Gràcies **Sílvia** per haver-me donat l'oportunitat de col·laborar amb el **PREC**, on sempre m'han rebut amb els braços oberts. Gràcies **Yolanda** per haver-me ensenyat tant d'estadística i de teatre. Y en especial, gracias **Delphine** por tu paciencia, enorme dedicación y por haberme guiado tanto.

Thanks **Marta** for the wonderful **Imperial College** experience, where I had the pleasure to compete with **Ulf** in the "accepted/rejected-paper challenge".

Gracias a los compañeros de **MCC-Spain** por todo vuestro apoyo. Thanks to all the **EPIC** colleagues for their valuable contributions to the papers. I, sobretot, gràcies a tots els participants d'aquests estudis que han fet possible aquesta tesi.

Finalment, gràcies a totes les persones meravelloses - **amics i família** - que m'envolten. Que per molts més anys pugui compartir nous cims (literals i vitals) amb tots vosaltres! Gràcies **Mario**, senzillament, per ser *casa*. I gràcies **papa** i **mama** per haver-me regalat la millor infància i totes les eines per arribar on em proposi. Ni amb mil tesis us podria agrair tot el que feu per mi. Per ara, us en dedico aquesta.

Moltíssimes gràcies a tots!

Marta

FUNDING

This doctoral thesis was supported by:

Consorcio de Inverstigación Biomédica de Epidemiología y Salud Pública (CIBERESP) through:

- a predoctoral contract.
- '*Ayudas para movilidad nacional'*, which supported a 2-week stay (from the 31st of October to the 11th of November 2016) at ISGLOBAL, under the supervision of Dr. Manolis Kogevinas.
- '*Ayudas para estancias breves en el extranjero para el Doctorado con mención internacional*', which supported a 3-months stay (from the 17th of September to the 17th of December 2018) at the School of Public Health, Imperial College London, under the supervision of Dr. Marta Blangiardo.

Universitat de Girona (UdG) through:

- MPCUdG2016 and GDRCompetUdG2017.
- '*Ajuts per a l'assistència a congressos internacionals de recerca*', which supported the attendance to the Annual InterLymph Meeting (25-28th of June 2018, Chicago, US).

LIST OF PUBLICATIONS

This thesis is presented as a compendium of three publications:

PAPER I

Title: Adherence to the Mediterranean diet and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition
Authors: Solans M, Benavente Y, Saez M, Agudo A, Naudin S, Saberi Hosnijeh F, Noh H, Freisling H, Ferrari P, Besson C, Mahamat-Saleh Y, Boutron-Ruault MC, Kühn T, Kaaks R, Boeing H, Lasheras C, Rodríguez-Barranco M, Amiano P, Huerta JM, Barricarte A, Schmidt J, Vineis P, Riboli E, Trichopoulou A, Bamia C, Peppa E, Masala G, Agnoli C, Tumino R, Sacerdote C, Panico S, Skeie G, Weiderpass E, Jerkeman M, Ericson U, Späth F, Nilsson LM, Dahm CC, Overvad K, Bolvig AK, Tjønneland A, de Sanjose S, Buckland G, Vermeulen, Nieters A, Casabonne D.
Journal: *International Journal of Cancer* 2018 Dec 26 [Epub ahead of print]
Impact Factor (2017): 7.360 (Q1 Oncology, position 23 of 223)
DOI: 10.1002/ijc.32091.

PAPER II

Title: Inflammatory potential of diet and risk of lymphoma in the European Prospective Investigation into Cancer and Nutrition
Authors: Solans M, Benavente Y, Saez M, Agudo A, Jakszyn P, Naudin S, Saberi Hosnijeh F, Gunter M, Huybrechts I, Ferrari P, Besson C, Mahamat-Saleh Y, Boutron-Ruault MC, Kühn T, Kaaks R, Boeing H, Lasheras C, Sánchez MJ, Amiano P, Chirlaque MD, Ardanaz E, Schmidt JA, Vineis PA, Riboli E, Trichopoulou A, Karakatsani A, Valanou E, Masala G, Agnoli C, Tumino R, Sacerdote C, Mattiello A, Skeie G, Weiderpass E, Jerkeman M, Dias JA, Späth F, Nilsson LM, Dahm CC, Overvad K, Petersen KEN, Tjønneland A, de Sanjose S, Vermeulen R, Nieters A, Casabonne D.
Journal: *European Journal of Nutrition* 2019 Mar 22 [Epub ahead of print]
Impact factor (2017): 4.423 (Q1 Nutrition and Dietetics, position 14 of 83)
DOI: 10.1007/s00394-019-01947-0

PAPER III

Title: Adherence to the Western, Prudent, and Mediterranean dietary patterns and chronic lymphocytic leukemia in the MCC-Spain study.
Authors: Solans M*, Castelló A*, Benavente Y, Marcos-Gragera R, Amiano P, Gracia-Lavedan E, Costas L, Robles C, Gonzalez-Barca E, de la Banda E, Alonso E, Aymerich M, Campo E, Dierssen-Sotos T, Fernández-Tardón G, Olmedo-Requena R, Gimeno E, Castaño-Vinyals G, Aragonés N, Kogevinas M, de Sanjose S, Pollán M*, Casabonne D*.
*Equal contribution
Journal: *Haematologica* 2018 Nov;103(11):1881-1888.
Impact Factor (2017): 9.090 (Q1 Hematology, position 5 of 71)
DOI: 10.3324/haematol.2018.192526.

Other related publications during the thesis period are listed in **Annex 1**.

LIST OF ABBREVIATIONS

AHEI, Alternate Healthy Eating Index
arMED, Adapted Relative Mediterranean Diet
ASR, Age-adjusted Incidence Rate
CI, Confidence Interval
CLL/SLL, Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma
CNPP, Center for Nutrition Policy and Promotion
CoDA, Compositional Data Analysis
DASH, Dietary Approaches to Stop Hypertension
DII, Dietary Inflammatory Index
DLBCL, Diffuse Large B-cell Lymphoma
EBV, Epstein-Barr Virus
EPIC, European Prospective Investigation into Cancer and Nutrition
FAB, American-British Classification
FFQ, Food Frequency Questionnaire
FL, Follicular Lymphoma
HEI, Healthy Eating Index
HIV, Human Immunodeficiency Virus
HL, Hodgkin Lymphoma
HLI, Healthy-lifestyle index
HR, Hazard Ratio
IARC, International Agency for Research on Cancer
ICD-O-2, International Classification of Diseases for Oncology, 2nd edition
ICD-O-3, International Classification of Diseases for Oncology, 3rd edition
InterLymph, International Lymphoma Epidemiology Consortium
ISD, Inflammatory Score of Diet
MCC-Spain, Multicase-control Spain
MM/PCN, Multiple Myeloma/Plasma Cell Neoplasm
NHL, Non-Hodgkin Lymphoma
NHLBI, National Heart, Lung, and Blood Institute
NIH, National Institutes of Health
NK, Natural Killer
OR, Odds Ratio
PCA, Principal Component Analysis
SD, Standard Deviation
rMED, Relative Mediterranean Diet
REAL, Revised European-American Lymphoma
RR, Relative Risk
USDA, United States Department of Agriculture
WCRF/AICR, World Cancer Research Fund/American Institute for Cancer Research
WHO, World Health Organization

LIST OF FIGURES, TABLES AND BOXES

List of figures

List of tables

List of boxes

*Original/adapted tables and figures were used with the permission of the copyright owners.
*Cover image: *Copyright © 2011 Michael Bonert (https://commons.wikimedia.org/wiki/User:Nephron), used under the CC BY-SA 3.0 license.*

ABSTRACT

Diet is a key modifiable risk factor for several neoplasms, but evidence for lymphoid malignancies is still inconsistent. Previous epidemiological studies were mostly focused on single dietary components (i.e. nutrients or food groups), while the role of overall diet has scarcely been studied. This thesis aims to explore the association of up to five dietary patterns and lymphoma risk, using data from two observational studies: the European prospective investigation into cancer and nutrition (EPIC) cohort and the Multicase-control Spain (MCC-Spain) study.

Results from the EPIC study showed that a greater adherence to the Mediterranean diet, measured through the adapted relative Mediterranean diet (arMED) score, was modestly associated with a decreased risk of overall lymphoma, but not with any specific histologic subtype. A more pro-inflammatory diet, quantified by the inflammatory score of diet (ISD), was modestly associated with mature B-cell non-Hodgkin lymphoma. Albeit with smaller numbers of cases, both the arMED and ISD showed suggestive (although statistically non-significant) associations with Hodgkin lymphoma. In the MCC-Spain study, we reported a positive lineal association between adherence to a Western dietary pattern and chronic lymphocytic leukemia, independently of Rai stage at diagnosis. By contrast, no associations were reported for a Mediterranean-like or Prudent dietary patterns.

Overall, our results suggest that dietary patterns may have a modest role in lymphoma etiology. This could partly explain the increasing lymphoma incidence trends during the second half of the 20th century, as well as higher incidence rates in westernized regions. These novel findings provide new insights into the possible link between modifiable lifestyle factors and lymphomagenesis, especially for Hodgkin lymphoma, which seems to be a subtype prone to be influenced by dietary exposures. Further large prospective studies with lymphoma subtype-specific data are warranted to confirm these findings.

RESUM

La dieta és un factors de risc modificable per nombrosos tipus de càncer però, pel què fa a les neoplàsies limfoides, l'evidència és encara inconsistent. Estudis epidemiològics previs al respecte s'han centrat majoritàriament en l'anàlisi de components de la dieta individuals (nutrients o grups d'aliments), mentre que el rol de la dieta en conjunt roman pràcticament inexplorat. Aquesta tesi avalua la relació entre cinc patrons de dieta i el risc de limfoma, mitjançant dades de dos estudis observacionals: *European Prospective Investigation into Cancer and Nutrition* (EPIC) i *Multicase-control Spain* (MCC-Spain).

En l'estudi EPIC, una major adherència a una dieta Mediterrània, mesurada mitjançant l'índex *adapted relative Mediterranean diet* (arMED), es va associar inversament amb el risc de limfoma, però no amb cap subtipus histològic. Una dieta pro-inflamatòria, quantificada a través de l'*inflammatory score of diet* (ISD), es va relacionar amb un augment modest del risc de neoplàsies de cèl·lules B madures. Tot i el reduït nombre de casos de limfoma de Hodgkin, resultats tant de l'arMED com l'ISD apunten a una possible associació (no estadísticament significativa) amb aquest tipus de limfoma. En l'estudi MCC-Spain vam reportar una associació lineal entre l'adherència a un patró de dieta occidental i la leucèmia limfàtica crònica, independentment de l'estadi Rai al diagnòstic. Per contra, no es van trobar associacions amb patrons de dieta saludables (mediterrani i prudent).

En conjunt, aquests resultats suggereixen que certs patrons de dieta estan modestament associats amb el risc de limfoma. Això contribuiria a explicar, en part, l'augment de casos de limfoma registrat durant la segona meitat del segle XX i les elevades taxes d'incidència que trobem en regions occidentalitzades. Aquestes troballes aporten noves evidències sobre la possible relació entre factors de risc modificables i limfomagènesi, en especial pel què fa al limfoma de Hodgkin, que sembla ser un subtipus de limfoma susceptible a aquest tipus d'exposicions. Futurs estudis prospectius amb informació detallada sobre els subtipus de limfoma contribuiran a confirmar-ho.

RESUMEN

La dieta es un factor de riesgo modificable para numerosos tipos de cáncer pero, en lo referente a neoplasias linfoides, la evidencia es aun inconsistente. Estudios epidemiológicos previos al respeto se han centrado en el análisis de componentes de la dieta individuales (nutrientes o grupos de alimentos), mientras que el rol de la dieta en conjunto yace prácticamente inexplorado. Esta tesis evalúa la relación entre cinco patrones de dieta y riesgo de linfoma mediante datos de dos estudios observacionales: *European Prospective Investigation into Cancer and Nutrition* (EPIC) y *Multicase-control Spain* (MCC-Spain).

En el estudio EPIC, una mayor adherencia a una dieta Mediterránea, evaluada mediante el índice "*adapted relative Mediterranean diet*" (arMED), se asoció inversamente con el riesgo de linfoma pero no con ningún subtipo histológico. Seguir una dieta pro-inflamatoria, cuantificada a través del *inflammatory score of diet* (ISD), se relacionó con un aumento modesto del riesgo de neoplasias de células B maduras. A pesar del reducido número de casos de linfoma de Hodgkin, resultados tanto de arMED como de ISD apuntan a una posible asociación (no estadísticamente significativa) con este tipo de linfoma. En el estudio MCC-Spain, reportamos una asociación lineal entre la adherencia a un patrón de dieta occidental y leucemia linfática crónica, independientemente del estadio Rai al diagnóstico. Por lo contrario, no se encontraron asociaciones con patrones de dieta saludables (mediterráneo y prudente).

En conjunto, estos resultados sugieren que ciertos patrones de dieta están modestamente asociados con el riesgo linfoma. Esto contribuiría a explicar, en parte, el aumento de casos de linfoma registrado durante la segunda mitad del siglo XX y las elevadas tasas de incidencia en regiones occidentalizadas. Estos hallazgos aportan nuevas evidencias sobre la posible relación entre factores de riesgo modificables y linfomagénesis, especialmente en lo que se refiere a linfoma de Hodgkin, que parece ser un subtipo de linfoma susceptible a este tipo de exposición. Futuros estudios prospectivos con información detallada sobre subtipos de linfoma contribuirán a confirmarlo.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Lymphoid neoplasms

1.1.1 Definition

Lymphoid neoplasms are a heterogeneous group of malignancies that arise from the malignant transformation of lymphoid cells at various stages of proliferation. The term encompasses both B-cell, T-cell, and natural killer(NK)-cell disorders.

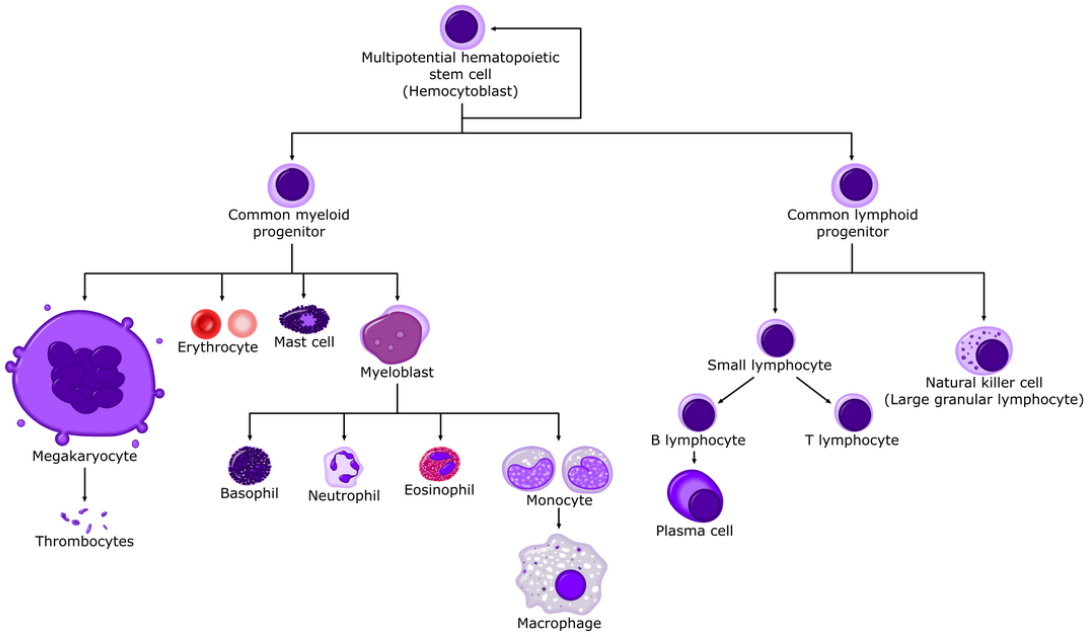


Figure 1. Diagram of hematopoiesis. Hematopoietic cells derive from pluripotent stem cells in the bone marrow and then differentiate into a range of cells either belonging to the myeloid or lymphoid lineage [original illustration by A. Rad., CC BY-SA 3.0].

1.1.2 Classification

Changes in our understanding of lymphoid neoplasms have resulted in the evolution of multiple classification schemes over the past 60 years[1,2]. Historically, classifications treated lymphomas and leukemias separately, based on whether the neoplastic proliferation was located in a lymph gland (e.g. lymph nodes) or proliferating cells circulated in blood, respectively. Initially, lymphomas were categorized based on morphology (Rappaport classification), morphology and prognosis (Working Formulation), or cell differentiation (Lukes and Collis or Kiel).

In 1994, the Revised European-American Lymphoma (REAL) classification, which further incorporated immunophenotypic, genotypic and clinical features, largely replaced the old classifications. One year later, these new definitions were introduced in the International Classification of Diseases for Oncology, 2nd Edition (ICD-O-2). On the other hand, leukemia was classified until 2000 by the French-American-British classification (FAB).

In 2001, the World Health Organization (WHO), based on the REAL and FAB schemes, introduced the current 'gold standard' for classifying all hematopoietic neoplasms[3]. Among the major changes, it replaced the artificial distinction between leukemia and lymphoma, and further included plasma cell neoplasms, often treated separately in previous schemes. Lately updated in 2008[4] and 2016[5], it represents a consensus classification for clinical and pathologic use which has been adopted worldwide, including the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)[6]. Overall, based on morphologic and immunologic characteristics, the WHO system distinguishes Hodgkin lymphoma (HL) from non-Hodgkin lymphoma (NHL). Cell lineage (B, T or NK), stage of differentiation (mature or immature) and additional morphologic, immunologic, genotypic, molecular, and clinical features are lately used to differentiate the various NHL subtypes. Broadly, NHL is categorized into T/NK-cell neoplasms and B-cell neoplasms. Among mature B-cell malignancies, the most frequent subtypes in the adult population are: diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), and multiple myeloma (MM).

Lack of consensus on how to translate prior classification schemes into the current classification, and uncertainty on how to group the more than 40 categories of subtypes, initially hampered the implementation of the WHO classification in epidemiological research. To overcome these limitations, first in 2007[2], and later in 2010[7], the International Lymphoma Epidemiology Consortium (InterLymph) proposed a hierarchical classification to group WHO lymphoma subtypes uniformly in epidemiologic subgroups (**Table 1**).

Table 1. Proposed 2008 WHO-based nested classification of malignant lymphoid neoplasms from the Pathology Working Group of the InterLymph[7].

1	2	3	4	5	6
LN	NHL	B-NHL	Mature B-NHL	***(See Table 1A)***	
			Precursor NHL	Precursor B-NHL	Precursor acute lymphoblastic leukemia/lymphoma B-cell, NOS
					Precursor B-lymphoblastic leukemia/lymphoma with recurrent genetic abnormalities (7 subtypes)
				Precursor (acute) lymphoblastic leukemia/lymphoma, unknown lineage	
		T/NK-NHL		Precursor T-NHL	Precursor T acute lymphoblastic leukemia/lymphoma
			Mature T/NK-NHL[a]	***(See Table 1B)***	
		NHL, NOS			
	HL	Classical HL	Lymphocyte rich HL		
			Mixed cellularity HL & lymphocyte-depleted HL	Mixed cellularity HL	
				Lymphocyte-depleted HL	
			Nodular sclerosis HL		
		Nodular lymphocyte-predominant HL			
	Composite HL and NHL				
	LPD	B-LPD	Mature B-LPD	Monoclonal B-cell lymphocytosis, CLL and non-CLL phenotypes	
				In situ mantle cell lymphoma	
				Monoclonal gammopathy of undetermined significance	
		T-LPD	Mature B-LPD	Lymphoid papulosis	
				Intraepithelial (intestinal) T-cell lymphoma-in situ	

LN, lymphoid neoplasm; NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; LPD, lymphoproliferative disease; ALK, anaplastic lymphoma kinase; EBV; Epstein Barr virus, HHV8, human herpesvirus 8.

[a] Categorized as T-cell or NK-cell when possible.

Table 1A. Hierarchical levels 5, 6 and 7 for mature B-NHL.

5	6	7
Chronic lymphocytic leukemia (CLL) & small lymphocytic lymphoma (SLL) & prolymphocytic leukemia B-cell & mantle cell lymphoma	CLL & SLL	CLL
		SLL
	Prolymphocytic leukemia B-cell	
	Mantle cell lymphoma	
Hairy cell leukemia		
Lymphoplasmacytic lymphoma (including Waldenström macroglobulinemia and gamma heavy chain disease) & marginal zone lymphoma (MZL)	Lymphoplasmacytic lymphoma	
	MZL	Extranodal MZL of mucosa-associated lymphoid tissue
		Nodal MZL (including its pediatric form)
		Splenic MZL
Splenic diffuse red pulp small B-cell lymphoma		
Hairy cell leukemia-variant		
Splenic B-cell lymphoma/leukemia, unclassifiable		
Plasma cell neoplasm	Plasmacytoma	Plasmacytoma bone
		Plasmacytoma extraosseous
	Plasma cell myeloma	
Follicular lymphoma	Pediatric follicular lymphoma	
	Primary cutaneous follicle Centre lymphoma	
Intrafollicular neoplasia 'in situ' follicular lymphoma		
Diffuse large B-cell lymphoma (DLBCL)	DLBCL, NOS	
	T-cell/histiocyte-rich large B-cell lymphoma	
	Primary DLBCL of the central nervous system	
	Primary cutaneous DLCBL, leg type	
	EBV+ DLCBL of the elderly	
	DLBCL associated with chronic inflammation	
	Intravascular large B-cell lymphoma	
	ALK+ large B-cell lymphoma	
	Plasmablastic lymphoma	
	Large B-cell (plasmablastic) lymphoma arising in HHV-8 associated multicentric Castleman's disease	
	Primary effusion lymphoma	
	Primary mediastinal (thymic) large B-cell lymphoma	
Lymphomatoid granulomatosis		
Burkitt lymphoma (including Burkitt leukemia)		
B-cell lymphoma, unclassifiable, with features intermediate between DLBCL and Burkitt lymphoma		
NHL B-cell, NOS		

Table 1B. Hierarchical levels 5, 6 and 7 for mature T/NK-NHL.

5	6	7
T-cell prolymphocytic leukemia		
T-cell large granular lymphocytic leukemia		
Chronic lymphoproliferative disorders of NK-cells		
Adult T-cell leukemia/lymphoma		
Aggressive NK-cell leukemia & extranodal NK/T-cell lymphoma, nasal type	Aggressive NK-cell leukemia	
	Extranodal NK/T-cell lymphoma, nasal type	
Peripheral T-cell lymphoma	Peripheral T-cell lymphoma, NOS	
	Angioimmunoblastic T-cell lymphoma	
	Hepatosplenic T-cell lymphoma	
	Enteropathy-associated T-cell lymphoma	
	Anaplastic large cell lymphoma, ALK-positive	
	Anaplastic large cell lymphoma, ALK-negative	
	Systemic EBV positive T-cell lymphoproliferative disease of childhood & hydroa vacciniforme-like cutaneous EBV positive T-cell lymphoproliferative disease of childhood	Systemic EBV positive T-cell lymphoproliferative disease of childhood
		Hydroa vacciniforme-like cutaneous EBV positive T-cell lymphoproliferative disease of childhood
	Subcutaneous panniculitis-like T-cell lymphoma	
	Primary cutaneous gamma-delta T-cell lymphoma	
	Primary cutaneous CD8 positive aggressive epidermotropic cytotoxic T-cell lymphoma	
	Primary cutaneous T-cell lymphoma, NOS	
	Primary cutaneous CD4 positive small/medium T-cell lymphoma	
Mycosis fungoides & Sézary syndrome & primary cutaneous CD30 positive lymphoproliferative disorders	Mycosis fungoides & Sézary syndrome	Mycosis fungoides
		Sézary syndrome
	Primary cutaneous CD30 positive lymphoproliferative disorders	Primary cutaneous anaplastic large cell lymphoma
		Lymphomatoid papulosis
NHL T-cell, NOS		

1.1.3 Incidence patterns

Together, lymphoid neoplasms comprise the eighth most common group of malignancies worldwide in both sexes[8]. Incidence rates are higher in men than in women for most subtypes, and generally increase with age, reaching their maximum at 75-99 years (with notable exceptions such as HL and Burkitt lymphoma)[9]. In addition, incidence rates show a marked geographical variability, being usually higher in industrialized countries compared to developing areas[8] (**Figure 1**).

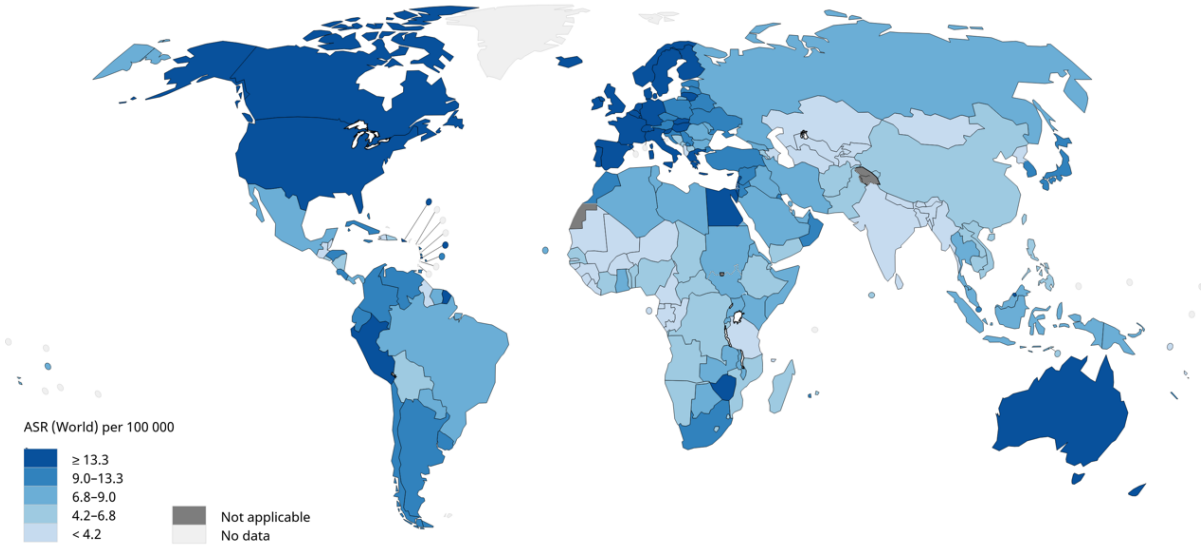


Figure 2. Estimated age-standardized incidence rates per 100,000 inhabitants of lymphoma (NHL, HL and MM combined) for both sexes in 2018 (Source: GLOBOCAN 2018[10]). Note that lymphoid leukemia (e.g. CLL/SLL) incidence rates are not provided in the figure, given that they are classified in the leukemia group together with myeloid neoplasms in GLOBOCAN datasets

Notably, the highest rates of lymphoid neoplasms are found in Israel/Lebanon, Australia, North America and Europe[8]. Within Europe, the age-adjusted incidence rate (ASR) of lymphoid neoplasms during 2000-2002 was 24.5 per 100.000 inhabitants, with higher rates in Southern Europe[9]. Among them, the commonest subtypes were plasma cell neoplasms (including MM) (ASR=4.6), CLL/SLL (ASR=3.8), DLBCL (ASR=3.1), and HL (ASR=2.4).

While HL incidence rates have remained stable over the years, NHL incidence rose dramatically in most Western countries throughout the second half of the 20th century and it have plateaued in the last decade[11,12]. Well-known risk factors, such as human immunodeficiency virus (HIV)[13], and changes in lymphoma diagnosis and registration may have contributed to this pattern, but do not completely explain this marked incidence increase[14].

1.1.4 Etiology

Lymphoid neoplasms' etiology is complex and multifactorial, with substantial heterogeneity among subtypes[15]. While some malignancies have been consistently linked to certain infections and severe immunosuppression, the etiology of most lymphoid neoplasms remains largely unknown.

1.1.4.1 Well-stablished risk factors

It has been long hypothesized that the striking bimodal age-incidence curve of HL, peaking in young adults (20-29 years) and in individuals aged over 60 years, may reflect etiological differences between these age groups[16]. There is strong evidence that Epstein Barr virus (EBV) may be involved in the causation of a great proportion of HL cases (namely younger cases), and clinical immune deficiency is a risk factor for few, but the cause of EBV-unassociated cases remains uncertain[17].

Congenital and acquired states of immunosuppression (mainly HIV infection) are the strongest factors known to increase NHL[18]. Other well-stablished risk factors include infectious agents (e.g. EBV, human T-cell lymphotrophic virus, human herpesvirus 8, and hepatitis C virus), autoimmune diseases (e.g. Sjögren's syndrome and systemic lupus erythematosus), and family history of blood malignancies[18,19]. However, not all risk factors are equally common within NHL subtypes, several associations are subtype-specific, as evidenced by the InterLymph NHL Subtypes Project[15] (**Figure 3**).

Exposure Category[A]	Specific Exposure	Prevalence (%) Cases	Cntls	P_{ASSET}	P_H	Overall NHL OR (95% CI)
Family history of hematologic malignancy[B]	Any	9.1	5.2	1.6×10^{-22}	3.5×10^{-2}	1.72 (1.54 - 1.93)
	NHL	4.0	2.0	1.7×10^{-13}	5.2×10^{-1}	1.79 (1.51 - 2.13)
	Leukemia	4.2	2.8	1.3×10^{-11}	3.9×10^{-5}	1.51 (1.29 - 1.77)
	Multiple myeloma	0.7	0.4	7.5×10^{-4}	2.2×10^{-2}	1.77 (1.15 - 2.72)
	Hodgkin lymphoma	1.1	0.6	2.0×10^{-3}	4.7×10^{-1}	1.65 (1.18 - 2.29)
Autoimmune disease[C]	Any B-cell activating disease	0.9	0.8	3.8×10^{-22}	9.8×10^{-10}	1.96 (1.60 - 2.40)
	Sjögren's syndrome	0.6	0.1	6.3×10^{-18}	7.3×10^{-9}	7.52 (3.68 - 15.4)
	Systemic lupus erythematosus	0.5	0.2	1.9×10^{-8}	1.8×10^{-1}	2.83 (1.82 - 4.41)
	Any T-cell activating disease	3.4	3.3	5.3×10^{-3}	1.2×10^{-2}	1.07 (0.95 - 1.21)
	Celiac disease	0.4	0.2	5.2×10^{-11}	5.1×10^{-8}	1.77 (1.05 - 2.99)
	Systemic sclerosis/scleroderma	0.1	0.1	5.1×10^{-3}	6.5×10^{-2}	1.03 (0.41 - 2.58)
HCV seropositivity[D]		2.3	2.2	2.3×10^{-8}	2.1×10^{-3}	1.81 (1.39 - 2.37)
Atopic disease[E]	Hay fever	18.2	20.1	9.1×10^{-9}	1.2×10^{-1}	0.82 (0.77 - 0.88)
	Eczema	9.8	9.8	5.0×10^{-5}	2.6×10^{-5}	1.01 (0.93 - 1.10)
	Allergy	22.0	24.4	5.9×10^{-5}	2.4×10^{-1}	0.86 (0.81 - 0.92)
Blood transfusion[F]	Transfusion occurring <1990	14.2	15.5	5.0×10^{-5}	1.3×10^{-2}	0.76 (0.67 - 0.87)
Anthropometric factors[G]	Body mass index as a young adult	21.1	17.9	4.2×10^{-9}	2.8×10^{-1}	1.95 (1.51 - 2.53)
	Height	53.2	52.0	1.7×10^{-3}	2.4×10^{-2}	1.20 (1.08 - 1.32)
Alcohol consumption (≥1 drink per month)	Any alcohol	69.3	72.1	8.9×10^{-8}	6.2×10^{-2}	0.87 (0.81 - 0.93)
	Wine	56.8	57.5	4.9×10^{-9}	1.4×10^{-2}	0.85 (0.79 - 0.91)
	Liquor	37.0	39.9	4.1×10^{-6}	6.6×10^{-1}	0.84 (0.78 - 0.91)
	Beer	44.9	47.2	9.3×10^{-4}	1.4×10^{-1}	0.90 (0.84 - 0.97)
Cigarette smoking[G]	Duration of smoking	57.0	56.7	2.2×10^{-5}	3.2×10^{-9}	1.06 (0.99 - 1.14)
Recreational sun exposure[G]		49.9	53.0	2.7×10^{-6}	7.9×10^{-1}	0.74 (0.66 - 0.83)
Socioeconomic status[G]		43.8	41.1	3.4×10^{-5}	6.1×10^{-2}	0.88 (0.83 - 0.93)
Occupational history[H]	Teacher	8.6	10.0	5.6×10^{-4}	6.2×10^{-3}	0.86 (0.77 - 0.95)
	Painter	2.0	1.8	4.8×10^{-3}	8.6×10^{-2}	1.22 (0.99 - 1.51)
	General farm worker	4.3	3.4	8.2×10^{-3}	3.4×10^{-1}	1.28 (1.10 - 1.50)

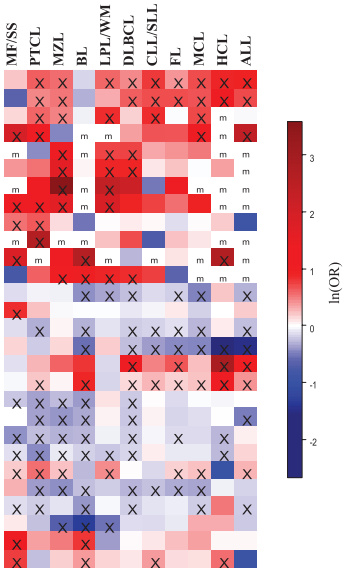


Figure 3. Risk factors associated with one or more NHL subtypes in the InterLymph NHL Subtypes Project (adapted from Morton *et al.* [15]). The columns list the exposure category, specific exposure, prevalence (all variables dichotomized) in cases and controls, p-value for association (P_{ASSET}), p-value for effect homogeneity (PH), and the odds ratio (OR) adjusted for age, race/ethnicity, sex, and study. For binary variables, the OR compares exposed vs unexposed, and for ordinal variables, OR compares highest vs lowest category. The colored grid indicates the log OR associated with the exposure for each subtype separately. Red and blue indicate that the exposure increases or decreases risk, respectively.

The role of other factors, such as lifestyle behaviors, in the etiology of lymphoid neoplasms, remains still uncertain. The following section summarizes the current state-of-the-art for dietary components and lymphoma risk.

1.1.4.2 Diet as a suggestive etiological factor

In 2007, the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) narratively reviewed the role of diet in the etiology of hematological neoplasms[20]. No conclusive statements were made but several suggestive associations were pointed out: i) vegetables, fruits, and alcoholic beverages were associated with a decreased risk of lymphoma, ii) meat and total fat with an increased risk of lymphoma, iii) and dairy products with an increased risk of NHL. In line with two other coetaneous reviews on diet and NHL[21,22], the Expert Panel remarked that more work into underlying mechanisms and further subtype-specific data would clarify the role of dietary factors in lymphomagenesis.

New data has started to accumulate during the last decade, mainly focused on the role of individual foods or nutrients in lymphomagenesis. No additional information has been provided in the recently released 2018 WCRF/AICR report[23], but recent meta-analyses have shed light on this relationship. Among them, the inverse link between alcohol consumption and NHL risk (namely for DLCBL and FL) is, to the date, one of the most consistent associations found[24], but lack of plausible mechanisms suggests caution in the interpretation of these results. Similarly, a meta-analysis on fats and NHL found that both high animal fat and total fat consumption increased the risk of overall NHL and DLBCL[25]. Cohort studies on specific micronutrients point to null associations between supplemented vitamins A, C and E, total vitamin D intake, as well as dietary lycopene intake, and risk of NHL[26]. Regarding large food groups, intakes of vegetables, vegetables and fruits combined, but not fruits alone, seem to be inversely associated with NHL[27,28], while positive associations have been reported with red meat[29] and dairy[29,30] intake and NHL risk.

Subtype-specific data is still limited, and mainly arises from studies on single nutrients or food groups. Taking CLL/SLL as an example, to the date, 9 prospective[31–39] and 13 case-control[40–52] studies have been published on this topic (**Table 2**). With the exception of few studies that found positive associations with consumption of processed meat and poultry[35], total carbohydrate[39] or fat (in women)[31] intake, and inverse associations with isoflavones consumption[36], generally large prospective studies found no associations between a wide range of dietary factors and CLL/SLL. By contrast, case-control studies have yielded contradictory results for meat[41–43,45,49], dairy products[42,43,45,47,52], fish[44,52] or vegetables and fruit[40–43,48,50] intake.

Overall, additional data is required to resolve inconsistent findings, especially where significant associations are limited to retrospective studies. Further studies within prospective cohorts with detailed dietary information and with a large number of cases to examine disease sub-type heterogeneity is warranted, pointing to a need for pooling and consortium efforts[21,22].

Table 2. Overview of epidemiological data on individual foods and nutrients and chronic lymphocytic leukemia risk

Author	Country	N	Fat	Protein	Red meat	Processed meat	Poultry	Egg	Dairy	Fish & seafood	Fruit & vegetables	Grains	Other
Meta-analysis[a] (n=8)													
Sergentanis *et al.* 2018[26]	-	5									NA		
Han *et al.* 2017[25]	-	4	NA										
Wang *et al.* 2016[30]	-	3							NA				
Caini *et al.* 2016[29]	-	9			NA	NA	NA		NA				
Yang *et al.* 2015[53]	-	5			NA								
Fallahzadeh *et al.* 2014[54]	-	7			NA								
Lu *et al.* 2014[55]	-	5							NA (Vit D)				
Chen *et al.* 2013[28]	-	10									NA		
Cohort (n=9)													
Bertrand *et al.* 2017[31]	US	253	♀										
McCullough *et al.* 2014[32]	US	267											NA (carbonated/sweet beverages)
Daniel *et al.* 2012[33]	US	979			NA	NA	NA			NA			
Kabat *et al.* 2012[34]	US	292											NA (Vitamins A, C, other antiox)
Rohrmann *et al.* 2011[35]	EU	234			NA			NA	NA				
Chang *et al.* 2011[36]	US	117											Isoflavones
Erber *et al.* 2010[37]	US	198											NA (Vitamin D)
Tsai *et al.* 2010[38]	US	1129	NA		NA	NA					NA		
Ross *et al.* 2002[39]	US	58	NA	NA	NA	NA	NA		NA	NA	NA	NA	Total carbohydrates
Case-control (n=13)													
Casabonne *et al.* 2016[40]	Spain	345											
Campagna *et al.* 2015[41]	Italy	~60			Beef, lamb		NA		NA		Citrus, tomatoes	Pasta	
Charbonneau *et al.* 2013[45]	US	218			NA	Hamburger	NA	NA	+/-				
Łuczyńska *et al.* 2013[46]	EU	161							25(OH)D				
Mikhak *et al.* 2012[47]	US	404							NA (Ca)				NA (Vitamins A and D)
Holtan *et al.* 2012[48]	US	218									Green leafy vegetables	NA	α-carotene
Aschebrook-Kilfoy *et al.* 2012[49]	US	25	NA	NA	NA	NA							
Chiu *et al.* 2011[50]	US	25									Citrus		
Koutros *et al.* 2008[51]	US	66											Vitamin B6
Chang *et al.* 2006[52]	Sweden	148	ω-3 FA						Ca	Fatty fish	Fiber		α -tocopherol, phosphorus, iron
Chang *et al.* 2005[42]	Sweden	147			NA				NA		Citrus ♂	NA	
Purdue *et al.* 2004[43]	Canada	174	NA	NA	Total meat	Total meat		NA	NA		Vegetables		Desserts
Fritschi *et al.* 2004[44]	Canada	58								NA			

N, number of cases (for meta-analyses, number of studies); US, United States; EU, Europe; Ca, calcium; NA, no association; ω-3 FA, omega-3 fatty acids; 25(OH)D, 25-hydroxyvitamin D. [a]All meta-analyses included both cohort and case-control studies, except from Sergentanis *et al.* which was based only on cohort studies. NA indicates non-associations found, while red and blue squares indicate statistically significant positive or inverse associations, respectively.

1.2 Dietary patterns

1.2.1 Background and definition

Traditionally, nutrition research focused on the study of single dietary factors (i.e. nutrients, foods or food groups). This reductionist approach has yielded numerous insights into the role of nutrition on health outcomes – especially when undernutrition and nutritional deficiencies were the prevailing diet – but poses several limitations[56–58]. First, intakes of many foods and nutrients are highly correlated, making it difficult to disentangle their separate effects. For instance, a diet rich in fruits and vegetables contains numerous potentially beneficial constituents (e.g. polyphenols, folate, vitamins, or fiber); therefore, studying only a specific component might result in misleading findings. Second, this reductionist approach often faults to consider substation effects, which are relevant in weight-stable populations, in which changes in one dietary component are accompanied by compensatory changes in others. Third, complex interactive, synergistic, and combined effects take place between foods and nutrients, and thus, such a focused approach may overlook the significance of overall diet.

The investigation of dietary patterns, which reflect the entire diet, has emerged as an alternative and complimentary approach. Dietary patterns are defined as the quantities, proportions, variety, or combination of different foods and drinks in diets, and the frequency with which they are habitually consumed[56]. Conceptually, they represent a broader picture of food and nutrient consumption, which account for inter-relations of food choices and represent the cumulative exposure to different dietary components. Overall, this multidimensional exposure seems to be more predictive of disease risk, as well as more easily translated into public health practice[57].

1.2.2 A priori and a posteriori methods

Two main approaches are commonly used to define dietary patterns: '*a priori*' and '*a posteriori*' methods. **Table 3** summarizes the main types of dietary pattern analysis, along with their strengths and limitations.

Table 3. Strengths and limitations of the main types of dietary pattern analysis (adapted from Shulze *et al.* 2018[59]).

A priori	***A posteriori***		
Indexes and scores	**Cluster analysis**	**PCA and factor analysis**	**Reduced rank regression**
Definition			
Hypothesis-oriented, based on previous scientific evidence.	Empirical approaches that use statistical methods to identify dietary profiles in a given community.		
Strengths			
Information on a variety of food items can be described by a single score.	Information on a variety of food items can be described by a few mutually exclusive clusters of people.	Information on a variety of food items can be described by a few underlying uncorrelated patterns.	Information on a variety of food items can be described by a few underlying uncorrelated patterns.
Easily reproducible and comparable.	Does not require prior theory; based only on the data.	Does not require prior theory; based only on the data.	Combines pathophysiological knowledge (hypothesis-oriented biomarkers) with study data (exploratory evaluation of food intake).
Particularly useful for evaluating associations between diet and disease endpoints.		Particularly useful for identifying existing patterns of food consumption.	Particularly useful for identifying patterns related to disease endpoints.
Limitations:			
Subjective selection of components and cut-offs.	Subjective decisions regarding cluster methods, numbers, distance measure.	Subjective decisions regarding number of patterns.	Subjective decisions regarding number of patterns.
Single components are considered as independent.	Descriptive analysis necessary to characterize patterns.	Unclear which food items characterize the pattern.	Unclear which food items characterize the pattern.
Dependent on strengths of evidence for hypothesis.	Procedure not related to outcomes.	Procedure not related to outcomes.	Dependent on knowledge and availability of response variables (e.g. disease biomarkers).
Assumes additive effects.		Only a low to moderate proportion of intake explained.	Only a low to moderate proportion of response variation explained.

PCA, principal component analysis

A priori or hypothesis-oriented methods define numerical indexes or scores on the basis of previous scientific evidence[57]. They are generally developed to quantify adherence to dietary-guidelines [e.g. the "Healthy Eating Index" (HEI)[60] and the "Dietary Approaches to Stop Hypertension index" (DASH)[61]] or designed reflect traditional or cultural diets, such as the Mediterranean diet[62]. Often, multiple indexes describe variations of the same dietary pattern (e.g. Mediterranean diet[63]), or use different scoring and weighting schemes, such as fixes cutoffs for recommended intakes or population-specific intakes. In all cases, the sum of all components included in the scoring do not reflect overall diet, but rather selected aspects of individual nutrition that may be related to health outcomes.

A posteriori or empirical approaches use statistical methods to identify dietary profiles in a given community[57]. They are further divided into principal component analysis (PCA) or factor analysis, cluster analysis, and reduce rank regression[64]. PCA examines the correlation matrix of food variables and searches for underlying traits (or factors) that explain the highest variability in the population's diet. Individuals receive a score for each of the factors derived, with higher values indicating higher adherence to each dietary pattern. In contrast, cluster analysis gathers homogenous subgroups or clusters of individuals with similar diets or the smallest differences in dietary variables within them. The patterns are then interpreted by analyzing the dietary profiles across clusters. Finally, reduced rank regression analysis is a mixture of the empirically and hypothesis-oriented approaches, as although it models dietary data, it also uses a priori knowledge to select the appropriate dietary variables.

1.2.3 Dietary patterns in cancer research: an overview

Examining the influence of overall diet has become a recent focus in cancer research. Indeed, the WCRF/AICR, which since 1997 has comprehensively reviewed and summarized evidence on individual foods and nutrients and their relationship with cancer risk, highlighted in its lately updated report that a more holistic focus might be more fruitful than a continuing search for specific dietary factors that might cause or protect against cancer[23]. Current evidence suggests that several *a priori*[65–68] and *a posteriori*[69] *derived* dietary patterns are consistently

associated with cancer risk across different populations, particularly with colon and breast cancers. Below, some of the most widely explored dietary patterns in cancer research are described in brief.

1.2.3.1 Mediterranean diet scores

Mediterranean diet scores, which quantify adherence to the traditional dietary patterns found in the olive-growing regions of the Mediterranean basin[70] (**Figure 4**), are among the most studied indexes worldwide. Whilst several variations exist[63], they all share the same key elements: a high consumption of fruits, nuts, seeds, vegetables, fish, legumes, cereals, and olive oil, a limited intake of meat and dairy, and a moderate consumption of alcohol[71].

The most frequently applied score is the Mediterranean diet score designed by Trichopoulou *et al.*[62], which has subsequently been adapted by the same or other authors. Among them, the relative Mediterranean diet score (rMED) from Buckland *et al.*[72], or its adaptation excluding the alcohol component, the adapted relative Mediterranean diet score (arMED)[73], have been extensively studied in the last decade. Whichever the score used to quantify adherence to the Mediterranean diet, there is a wealth of evidence supporting its role in increasing longevity and preventing many non-communicable chronic diseases, such as cardiovascular diseases or several types of cancer[74]. Concretely for cancer prevention, a recent meta-analysis[65] including 2 randomized trials, 51 cohort and 30 case-controls studies, recently reported that the highest adherence to a Mediterranean diet score was inversely associated with risk of colorectal and breast cancers. Sustained by less studies (or mostly arising from case-control studies), an association with gastric, liver, head and neck, and prostate cancers was also reported.

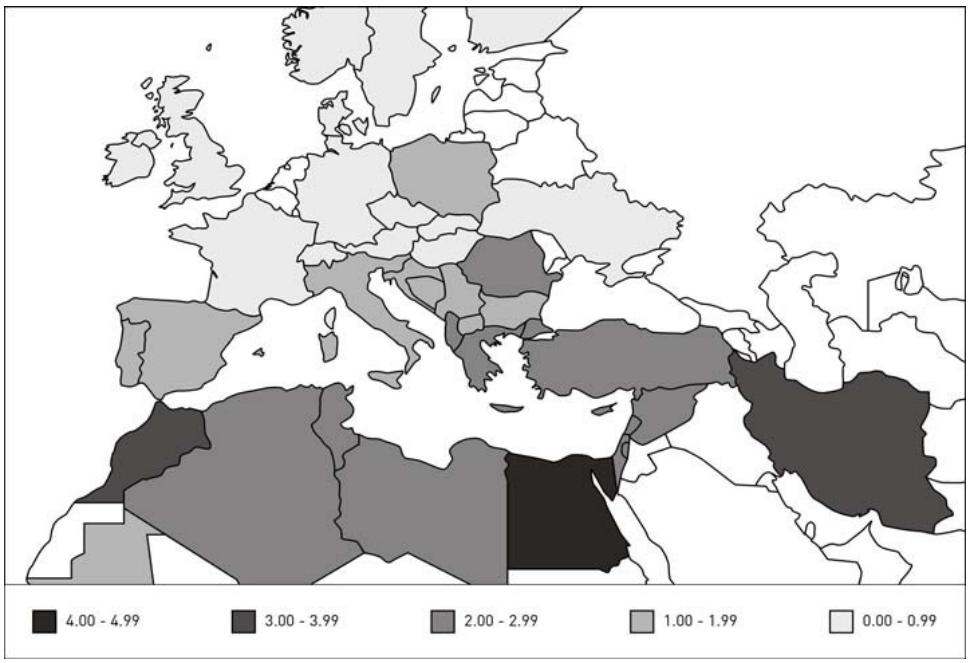


Figure 4. Map of the adherence to the Mediterranean dietary pattern in the period 2000-2003 (adapted from da Silva *et al.* 2009[75]). Note that it was measured using the Mediterranean Adequacy Index (higher values indicate higher adherence to the dietary pattern).

1.2.3.2 Scores based on dietary recommendations

Several diet quality indexes have been largely evaluated in terms of cancer prevention. Among them, the HEI[60,76–78], the Alternative Healthy Eating Index (AHEI)[79,80], and the DASH[61] scores are distinct scores of major public health importance, notably in the United States. The original HEI[76] was released by the United States Department of Agriculture (USDA) Center for Nutrition Policy and Promotion (CNPP) in 1995, to measure how well diets conform to federal dietary guidance. The score has been subsequently updated into the HEI-2005[77], HEI-2010[78], and HEI-2015[60] to reflect the changes in the Dietary Guidelines for Americans[81]. Based on the original HEI, McCullough *et al*. developed in 2002 the AHEI[79], which only included foods and nutrients predictive of chronic disease, and which was later updated in 2012[80]. Finally, the DASH score[61], promoted by the by the National Heart, Lung, and Blood Institute (NHLBI), is one of the most well-known dietary strategies for lowering blood pressure. No single score[82–84] is considered to optimally reflect the DASH diet; however, to the date, the DASH score established by Fung *et al.*[83] is the most frequently used in the literature.

Overall, the three scoring systems share key components that are, in turn, similar to those present in the Mediterranean diet score[59] (**Table 4**).

Table 4. Key components of the Mediterranean diet, DASH, AHEI, and HEI-2015 scores.

Components	**MD**[62,72]	**DASH**[61]	**AHEI**[80]	**HEI-2015**[60]
Cereals		Whole grain	Whole grain	Whole grain
				Refined grain
Vegetables				
Fruits				Especially whole
Legumes, nuts, seeds				
Fish				
Meat	Red and processed meat	Fatty meat	Red and processed meat	Included in the total protein component
Dairy products		Low fat	-	Low fat or fat-free
Fats	Olive oil or ratio MUFA/saturated FA		Polyunsaturated and long-chain Ω-3 fatty acids	PUFA + MUFA
			Trans fat	Saturated FA
Sweets/sweetened beverages	-			
Alcohol		-		
Sodium	-			

MD, Mediterranean diet; DASH, dietary approaches to stop hypertension; AHEI, adapted healthy eating index; HEI-2015, healthy eating index (2015 version); PUFA, polyunsaturated fatty acids; MUFA, monounsaturated fatty acids; FA, fatty acids.
Blue, yellow, and red indicate encouraged, in moderation or discouraged intake.

Overall, diet quality indexes have been associated with lower risk of cancer in a recent cohort meta-analysis[66]. Concretely, diets that scored highly on the HEI, AHEI, and DASH indexes were associated with a reduced risk of overall cancer incidence or mortality. Subgroup analysis revealed an inverse association between the highest level of diet quality and risk of colorectal, esophageal, lung, gallbladder, pancreatic, prostate, head/neck, and hepatocellular carcinoma.

Nutrition-based cancer prevention guidelines have been also operationalized into a score and largely evaluated in epidemiological studies. Concretely, in 2007, the WCRF/AICR formulated a series of recommendations for cancer prevention[20], based on a meta-analysis of the most comprehensive collection of published evidence at that time on diet, weight, and physical activity in association with cancer. A scoring system was then developed to assess the degree of adherence to such recommendations in the European Prospective Investigation into Cancer and Nutrition (EPIC) study[85] and later applied in different epidemiological studies. Substantial evidence indicates that adherence to the 2007 WCRF/AICR score is associated to a lower risk of cancer incidence, mostly breast cancer[68]. Ten year later, and based on an updated literature review, a new version of the recommendations has been released[23]. The current recommendations include: i) be healthy weight, ii) be physically active, iii) eat a diet rich in whole grains, vegetables, fruits and beans, iv) limit the consumption of 'fast foods' and other processed foods high in fat, starches or sugars, v) limit consumption of red and processed meat, vi) limit consumption of sugar sweetened drinks, vii) limit alcohol consumption, viii) do not use supplements for cancer prevention, ix) for mothers, breastfeed your baby, if you can, and x) after a cancer diagnosis: follow our recommendations if you can. The impact of the current guideless in the prevention of cancer remains yet unexplored.

1.2.3.3 Scores assessing the inflammatory potential of diet

Due to the role of inflammatory pathways in the pathogenesis and progression of several neoplasms[86], it has been long suggested that the association between diet and cancer may be in part mediated through diet-induced inflammation[23]. Over the few last years, several scores on the inflammatory potential of diet have been used to address such question. Among them, the dietary inflammatory index (DII)[87], and its adaptation into the inflammatory score of diet (ISD)[88] in the EPIC study, are the most widely used. They are both literature-derived tools designed to reflect all evidence available linking foods/nutrients intake with inflammatory markers, ultimately categorizing individuals' diets from maximally anti-inflammatory to

maximally pro-inflammatory. The scoring system and the main differences between the DII and ISD are further detailed in **Box 1.**

Box 1| Operationalization of the DII and ISD

- **A review of articles** published from 1950 to 2010 was performed, resulting in 1943 studies linking to a total of 45 food parameters with inflammatory biomarkers (IL-1β, IL-4, IL-6, IL-10, TNF-α and CRP)[87]. The 45 components include carbohydrate, protein, fiber, fats, vitamins, minerals, flavonoids, ethanol, total energy, and several foods – garlic, ginger, onion, tea, caffeine, and spices.

- **A score for each food parameter** was calculated giving +1, -1, or 0 to each article if the effects were pro-inflammatory, anti-inflammatory or absent, respectively[87]. Each score was weighted according to the study design: 10 (experimental design), 8 (observational), 7 (case-control), 6(cross-sectional), 5 (experimental with animals), 3 (cell culture), and finally summed into an overall inflammatory score for each component.

- **Operationalization of the DII**[87] **–** Each dietary component intake is standardized using a regional worldwide database as referent, resulting in z-scores which are later converted to centered-percentiles. Each standardized intake is multiplied is by its respective inflammatory effect score. All 'food parameter-specific DII scores' are summed to finally obtain the overall DII.

- **Differences between the DII and ISD** – The ISD is an adaptation of the DII that has been developed in the EPIC study[88]. It differs from the DII in three aspects:
    - Food intakes are standardized using the EPIC population as a reference population.
    - The anti-inflammatory weight of alcohol is restricted to moderate consumption (less than 30-40 ethanol g/day).
    - Total fats are dismissed to avoid overestimation of their inflammatory effect, which is likely to be already represented by separate components of fats.

During the past several years, literature linking the DII and cancer has grown exponentially. Recent meta-analyses[67,89,90] analysing such studies report consistent positive associations between higher DII and cancer incidence across different cancer subtypes, study populations and study designs.

1.2.3.4 Empirically derived dietary patterns

Empirically derived dietary patterns have been also related to overall and site-specific cancer risk. Recently, Grosso *et al.*[69] performed a systematic review of a posteriori-derived dietary patterns grouped as "healthy" (fruit and vegetable-based diets, named 'healthy' or 'prudent', and some traditional patterns, such as the 'Medirerranean-like') or "unhealthy" (characterised by, but not limited to, red and processed meat, sugary drinks and salty snacks, starchy foods, and refined carbohydrates, and generally referred as "Western-like"). Results from prospective studies supported an inverse association between healthy dietary patterns and colorectal, breast, and lung cancers, and a positive association between unhealthy dietary patterns and colorectal cancer risk. The risk of cancer at several other sites was also associated with dietary patterns, but evidence was mainly driven by case–control studies.

1.2.4 Dietary patterns and lymphoid neoplasms risk

In contrast to solid neoplasms, there is few evidence on dietary patterns and lymphoid neoplasms. To the best of our knowledge, one cohort[91] and five case-control studies[41,92–95] have been published on this topic. Their design and main findings are summarized in **Table 5**.

Whilst the prospective study of Erber *et al.*[91] did not find consistent associations with empirically derived patterns and NHL (with the exception of an inverse association between a Vegetables dietary pattern and NHL among Caucasian women), two case-control studies found that a pattern rich in 'Desserts and sweets'[92] or in 'Meat, fat and sweets'[93], were associated with overall HL or NHL risk, respectively. As far as a priori indexes are concerned, a case-control study in the Mediterranean island of Sardinia did not find an association between

adherence a Mediterranean diet and NHL[41]. The exposure variable was a variation of the rMED score, but differed substantially in key elements: olive oil and dairy were not included in the scoring, the meat component did not consider white meat, and only wine consumption, not other beverages, computed in the alcohol component. Finally, two Italian case-control studies have recently reported an association between the inflammatory potential of diet, by means of the DII, with overall NHL and DLBCL[94], but not with HL[95].

Table 5. Summary of results from epidemiological studies on dietary patterns and lymphoid neoplasms.

Author	Country	Controls	Cases	Approach	Dietary pattern	Results
Cohort						HR (95% CI)
Erber *et al.* (2009)[91]	US	193,050 (total subjects)	939 NHL Subtype analysis: DLBCL, FL and CLL/SLL	Factor analysis	Vegetables	♂: 0.96 (0.73; 1.26) ♀ :0.83 (0.61; 1.13) ♀ Caucasian: **0.56 (0.33; 0.95)**
					Fat and Meat	♂:1.11 (0.78; 1.57) ♀: 0.85 (0.58; 1.24) FL men: **5.16 (1.33; 20.00)**
					Fruit and Milk	♂: 0.88 (0.66; 1.17) ♀: 1.04 (0.76; 1.44) ♀ SLL/CLL: **0.16(0.05;0.44)**
Case-Control						OR (95% CI)
Epstein *et al.* (2015)[92]	US	563	435 classical HL Subtypes: nodular sclerosis and multiple cellularity	PCA	High vegetable	0.78 (0.52; 1.16)
					Western-Style	Total: 1.39 (0.86; 2.26) **≥50 years: 3.34 (1.02; 10.91)**
					Fruit/low-fat dairy	0.70 (0.45; 1.07)
					Desserts/sweets	**1.50 (1.02; 2.21)**
Ollberding *et al.* (2014)[93]	US	460	336 NHL Subtypes: B-cell, FL, CLL/SLL, MZL, other B-cell, T-cell	Factor analysis	Meat, fat and sweets	Total: **3.6 (1.9; 6.8)** FL: **3.1 (1.2; 8.0)** DLBCL: **3.2 (1.1; 9.0)** MZL: **8.2 (1.3; 51.2)**
					Vegetables and fruit	0.9 (0.6; 1.4)
Campagna *et al.* (2015)[41]	Italy	446	322 lymphoma Subtypes: B-cell, DLBCL, CLL, FL, MM, HL	*A priori*	Mediterranean diet score (adaptation of rMED score)	Total: 0.9 (0.6; 1.5) DLBCL: 0.4 (0.1; 1.0)
Shivappa *et al.* (2017)[94]	Italy	984	536 NHL Subtypes: DLCBL, FL	*A priori*	Dietary inflammatory index	Total: **1.61 (1.07; 2.43)** Males: **2.14 (1.25; 3.67)** DLBCL: **1.84 (1.09; 3.10)**
Shivappa *et al*. (2018)[95]	Italy	186	179 HL	*A priori*	Dietary inflammatory index	1.20 (0.71; 2.04)

HR, hazards; RR, risk ratio; OR, odds ratio; CI, confidence interval; n/a, not applicable; NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; CLL/SLL, chronic lymphocytic leukemia/small lymphocytic lymphoma; MM, multiple myeloma; MZL, marginal zone lymphoma; PCA, principal component analysis; rMED, relative Mediterranean diet score. All results comparing highest x-tile with lowest x-tile. Results in bold are statistically significant at p-value<0.05 level.

2. RATIONALE

Lymphoid neoplasms' incidence shows a marked geographical variability, with most with most cases occurring in western countries[8,96]. In addition, these patterns vary with migration and nativity, pointing to an influence of acculturation on lymphomagenesis[97,98]. Overall, the parallel between lymphoma incidence rates and the level of westernization, together with the dramatic rise of NHL incidence since 1970's until 2000, suggests a potential role of environmental factors in the etiology of lymphoma.

While individual roles of lifestyle behaviors (e.g. physical activity, body mass index, smoking, alcohol consumption or diet) have been extensively investigated for solid neoplasms[23], evidence on lymphoma is scarcer and still inconsistent. For diet in particular, most data arise from studies on nutrients or individual food components, which, to the date, have reported inconsistent findings and not generated strong evidence[20–22].

Focusing on overall dietary patterns better captures cumulative and interactive effects between dietary factors, and may thus be more predictive of disease risk than individual foods or nutrients. Indeed, in the recently released 2018 WCRF/AICR report[23], the panel emphasized the importance of adopting a more holistic focus in cancer prevention research, by considering how different patterns of diet and physical activity combine to create a metabolic state prone to be associated with cancer development, rather than focusing on the singular effects of specific dietary factors.

The influence of numerous *priori*[65–68] and *a posteriori*[69] dietary patterns in the etiology of solid neoplasms has been extensively studied. To the date, however, there are few published data on lymphoid neoplasms[41,91–95], and generally unpowered to find subtype-specific associations. Deeping into the study of dietary patterns in lymphoid neoplasms could shed a light on the association between diet and lymphomagenesis.

3. OBJECTIVES

This thesis aims to evaluate the influence of several scores and dietary patterns in the etiology of lymphoid neoplasms. To address this general aim, the following specific objectives were defined:

1. To prospectively investigate the association between adherence to a Mediterranean diet, estimated using the arMED score, and the risk of primary incident lymphoma and its subtypes (HL, NHL, mature T/NK-cell neoplasms, mature B-cell neoplasms, DLCBL, FL, CLL/SLL, MM/plasma cell neoplasms (PCN), and other B-cell neoplasms) in the EPIC study.

2. To prospectively investigate the association between the inflammatory potential of diet, measured using the ISD, and the risk of primary incident lymphoma and its subtypes (HL, NHL, mature T/NK-cell neoplasms, mature B-cell neoplasms, DLCBL, FL, CLL/SLL, MM/PCN and other B-cell neoplasms) in the EPIC study.

3. To analyze the association between three *a posteriori* dietary patterns (Mediterranean-like, Prudent and Western), and CLL/SLL, overall and by Rai stage, in the multicase-control Spain (MCC-Spain) study.

4. METHODS

This thesis is based on data from two observational studies: the EPIC cohort (for objectives 1 and 2) and the MCC-Spain study (for objective 3). **Table 6** provides an overview of the study design, the dietary patterns' construction, and the statistical analyses performed, which are further detailed in the following section.

Table 6. Overview of methodology.

Objective	**1:** To assess the association between adherence to the Mediterranean diet and risk of lymphoma and its subtypes	**2**: To assess the association between the inflammatory potential of diet and risk of lymphoma and its subtypes	**3:** To analyze the association between *a* Mediterranean-like, Prudent and Western dietary patterns and CLL/SLL
Study	EPIC cohort	EPIC cohort	MCC-Spain study
Design	Prospective cohort (23 centers in 10 European countries)	Prospective cohort (23 centers in 10 European countries)	Multicentric case-control study (11 centers from 5 Spanish areas)
Cases[a]	3,136 lymphoma, categorized into major subtypes[b]	3,136 lymphoma, categorized into major subtypes[b]	369 CLL/SLL
Controls or participants[a]	476,160 participants	476,160 participants	1,605 controls
Exposure	Adapted relative Mediterranean diet (arMED) score	Inflammatory score of diet (ISD)	Western/Prudent/ Mediterranean dietary patterns extracted by PCA in the EpiGEICAM study[c].
Statistical models	Cox proportional hazards regression	Cox proportional hazards regression	Mixed logistic regression
Statistical software	Stata 14.1	Stata 14.1	Stata 14.1

[a]After exclusions for the specific analyses.
[b]Lymphoma cases were classified into HL and NHL. Among NHL, into mature T/NK-cell and mature B-cell neoplasms. Among the mature B-cell neoplasms, into DLCBL, FL, CLL/SLL, MM/PCN, and other B-cell neoplasms.
[c]The EPIGEICAM study is a multicentric case-control study on female breast cancer in Spain[99].

4.1 The EPIC study

4.1.1 Study population

EPIC is a large prospective cohort study designed to investigate the relationship between diet, lifestyle, environmental factors and cancer[100,101]. 521,324 subjects, mostly aged 30 to 70 years, were recruited between 1992-2000 in 23 centers from ten European countries (Denmark, France, Germany, Greece, the Nederland's, Italy, Norway, United Kingdom, Spain and Sweden) (**Figure 5**). The ethical review boards from the International Agency for Research on Cancer (IARC) and all local participating centers approved the study, and all participants gave their informed consent.

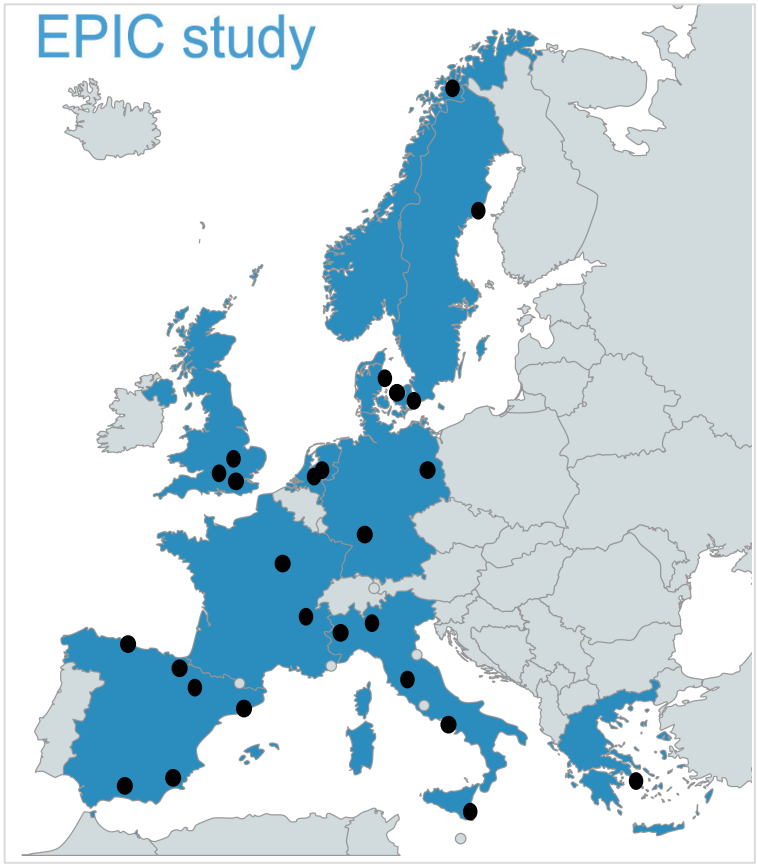


Figure 5: Countries (in blue) involved in the EPIC study.

For the current analyses, we excluded prevalent cancer cases (n=25,184), subjects with missing follow-up information (n=4,148), with incomplete or without dietary

information (n= 6,259), or those in the highest and lowest 1% of the distribution for the ratio of energy intake to estimate energy requirement (n=9,573). Therefore, our analyses were based on 476,160 subjects among whom 3,136 incident lymphoma cases occurred.

4.1.2 Data collection

Validated country-specific questionnaires were used to record the usual diet during the previous year[101,102]; namely through quantitative or semi-quantitative food frequency questionnaire (FFQ), administered through a personal interview or self-administered. Lifestyle questionnaires were used to obtain information on sociodemographic characteristics, physical activity, reproductive history, use of oral contraceptives and hormone replacement therapy, medical history and alcohol and tobacco consumption. Anthropometric measures were also ascertained at recruitment.

4.1.3 Follow-up and outcome assessment

Incident lymphoma cancer cases were identified by population cancer registries for Denmark, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom. A combination of methods was used in France, Germany and Greece, as detailed previously[101]. Mortality data were also obtained from regional or national mortality registries. The follow-up period was defined from the age at recruitment to the age at first cancer diagnosis, death or last complete follow-up, whichever occurred first. Censoring dates for the last complete follow-up ranged from June 2008 to December 2013, depending on the EPIC center.

Lymphoma cases, initially registered using the ICD-O-2, were reclassified according to the ICD-O-3 and categorized using the InterLymph Pathology Working Group classification, which is based in the 2008 WHO classification[7] (see **Figure 6**). Overall, during an average follow-up of 13.9 years, 3,136 cases were diagnosed.

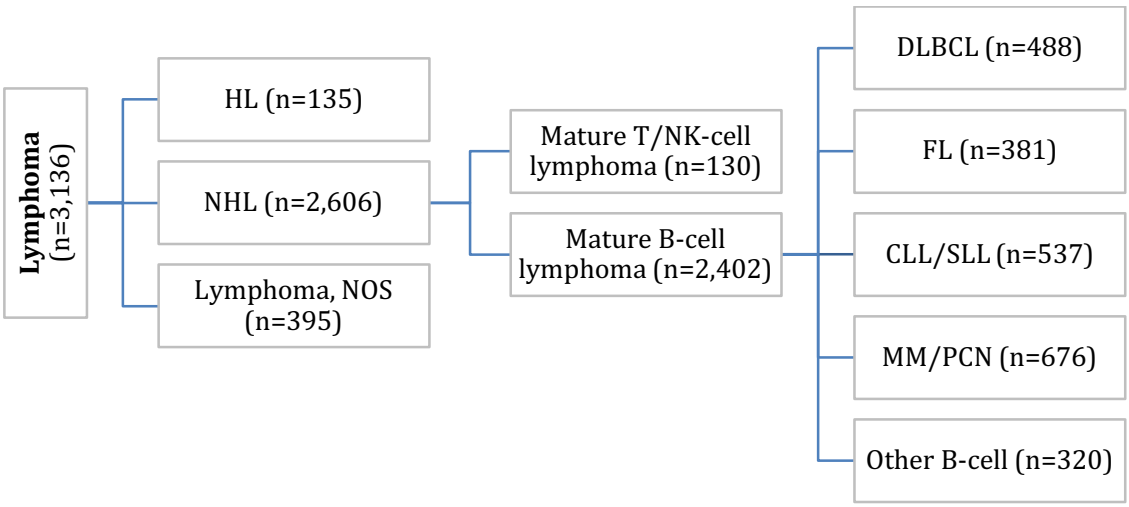


Figure 6: Lymphoma major subgroups included in the current EPIC analyses. Note that the n indicates the number of cases after exclusion for the specific analyses. Among overall NHL, there were 37 precursors of NHL and 37 individuals without B- or T- cell information. Other B-cell category included those cases in which the B-cell lymphoma subtype is unknown or does not fall within the above mentioned subtypes.

4.1.4 Exposure assessment

4.1.4.1 The arMED score

The level of adherence to the Mediterranean diet was assessed using the arMED score[73], which excludes alcoholic beverages as they have been inversely associated with several lymphoma subtypes[24]. The arMED is a 16-point linear score that incorporates eight key dietary components: six components presumed to reflect the Mediterranean diet (fruit, vegetables, legumes, fish, olive oil and cereals) and two components consumed in low quantity in the Mediterranean diet (dairy products and meat). Intake of each component was calculated as a function of energy density (g/day/1000kcal) and divided into tertiles (estimated using the overall study population). A value of 0 to 2 was assigned for the first, second, and third tertile of intakes for the components, while the scoring was inverted for the components presumed to not fit the Mediterranean diet. The scoring for olive oil was adapted owing to the low consumption of non-Mediterranean countries, by assigning 0 to nonconsumers, 1 for subjects below the median and 2 for subjects equal or above this median (the median was calculated using the overall study population and considering only consumers) (**Table 7**). The resulting points were

finally summed to define the arMED score, that ranged from 0 to 16 (from the lowest to the highest adherence).

Table 7. Construction of the arMED score

arMED score components[a] (g/1000kcal/day)	**Groups of intakes and scoring assigned**		
	0	1	2
Vegetables (excluding potatoes)	Tertile 1	Tertile 2	Tertile 3
Fruit (including nuts and seeds)	Tertile 1	Tertile 2	Tertile 3
Legumes	Tertile 1	Tertile 2	Tertile 3
Fish and seafood	Tertile 1	Tertile 2	Tertile 3
Cereals	Tertile 1	Tertile 2	Tertile 3
Olive oil	Non-consumers	<Median[b]	>Median[b]
Meat	Tertile 3	Tertile 2	Tertile 1
Dairy	Tertile 3	Tertile 2	Tertile 1

[a]The rMED score also includes alcohol, traditionally drunk in moderation in the Mediterranean diet, score as a dichotomous variable: two points are assigned to moderate consumers (5-25g/day for women and 10-50g/day for men) and 0 to those below or beyond the range.
[b]Median calculated only within consumers

4.1.4.1 The ISD

The inflammatory potential of diet was assessed using the ISD[88]. 28 dietary components available in the EPIC database for all centers were selected (**Table 8**). The intake of each food parameter was standardized using the mean and standard deviation of the EPIC population (**Table 8**). The obtained z-scores were then converted to percentile scores to avoid the right skewness of data, and then centred on 0 by doubling each percentile score and subtracting 1. The centred percentile values were then multiplied by its respective inflammatory weight (**Table 8**). The food parameter-specific inflammatory score was then summed to obtain the overall ISD for each individual. Overall, the ISD is a relative index that allows categorizing individuals' diets on a continuum from maximally anti-inflammatory (corresponding to lower scores) to maximally pro-inflammatory (higher scores).

Table 8. Food parameters, inflammatory effect scores (weights), and mean and standard deviation of the intake in the EPIC study population used to calculate the ISD.

Dietary parameters[a]	Inflammatory effect score[87]	Daily mean intake[b]	SD[b]
Macronutrients			
Carbohydrate (g)	0.097	228.06	74.40
Protein (g)	0.021	86.70	27.41
Fiber (g)	-0.663	22.78	7.72
Fats (components)			
Saturated fat (g)	0.373	31.17	12.95
MUFA (g)	-0.009	30.03	13.51
PUFA (g)	-0.337	13.46	6.18
Cholesterol (mg)	0.110	315.08	150.79
Alcohol			
Ethanol (g)	-0.278[c]	11.96	17.11
Vitamins			
Vitamin A (RE)	-0.401	830.43	745.46
β-carotene (µg)	-0.584	3,577.36	2,847.03
Thiamin (mg)	-0.098	1.34	0.50
Riboflavin (mg)	-0.068	1.86	0.75
Vitamin B6 (mg)	-0.365	1.87	0.62
Folic acid (µg)	-0.190	308.51	119.81
Vitamin B12 (µg)	0.106	6.51	4.02
Vitamin C (mg)	-0.424	127.26	69.44
Vitamin D (µg)	-0.446	4.16	3.45
Vitamin E (mg)	-0.419	12.17	5.77
Minerals			
Fe (mg)	0.032	13.06	4.27
Mg (mg)	-0.484	359.87	110.78
Flavonoids			
Flavan-3-ol (mg)	-0.415	118.10	159.07
Flavones (mg)	-0.616	11.52	8.61
Flavonols (mg)	-0.467	39.67	33.47
Flavonones (mg)	-0.250	39.55	45.29
Anthocyanidins (mg)	-0.131	39.72	54.10
Isoflavones (mg)	-0.593	1.73	9.13
Foods			
Onion (g)	-0.301	11.74	13.46
Other			
Total energy (kcal)	0.180	2,075.07	619.22

SD, standard deviation; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids.
[a]Based on dietary information (did not consider supplements)
[b]Obtained from EPIC study population
[c]Only up to consumers (<40g/day), for heavy consumers the inflammatory effect score was 0, owing its U/J-shape association with inflammatory biomarkers[103–105].

4.1.5 Statistical analyses

Cox proportional hazard models were used to estimate the hazard ratio (HR) and 95% confidence intervals (CI) of the association between the scores and lymphoma risk. The arMED score was analyzed both as a continuous variable (per 1-unit increase) and as a categorical variable [in low (0-5), medium (6-9) and high (10-16) categories, as described previously[73]]. Similarly, the ISD was analyzed as a continuous variable (per 1-standard deviation (SD) increase) and as a categorical variable (in quartiles). In addition, Cox models were fitted with the ordinal variables as continuous to test for linear trend for comparison with published literature on solid cancers.

Two models with two levels of adjustment were used: a basic model, stratified by center, sex and age at recruitment (in 1-year categories), and a multivariable model, further adjusted for body mass index (BMI) (<25, 25-30, ≥30 kg/m^2), total energy intake (continuous, kcal/day), educational level (no formal education, primary school, secondary school, technical or professional training, University, unknown [3.6%]), height (continuous, cm), physical activity level based on the Cambridge physical activity index (inactive, moderately inactive, moderately active, active, unknown [1.9%]), smoking status (never, former, current and, unknown [2.0%]), and alcohol intake at recruitment (continuous, g/day).

We tested for interaction by age, sex, alcohol intake, and smoking (and BMI for the ISD, as well) with the continuous scores using a likelihood ratio test. In addition, sensitivity analyses were performed by repeating main Cox analyses:

- excluding cases diagnosed within the first two years of follow-up (n=259 cases), to account for potential reverse causality.
- excluding participants without complete data (n=226 cases).
- restricting HL analysis to classical HL, which may have a different etiology[7].
- excluding alcohol from the ISD, to confirm it was not the only element driving the association.
- using the rMED score[106], which also considers alcohol. Alcohol was scored as a dichotomous variable, assigning two points for moderate consumers (5–25 g/day for women and 10–50 g/day for men) and 0 points for those

above and below the sex-specific range, owing its beneficial effects if consumed in moderation.

Schoenfeld residuals were used to ensure that the proportional hazard assumption was met in all models. Two-sided p-values were reported with statistical significance set at $p<0.05$. All analyses were performed by using STATA statistical software, version 14 (Stata Corporation, College Station, Texas).

4.2 The MCC-Spain study

4.2.1 Study population

The MCC-Spain is a multicentric case–control study evaluating environmental and genetic factors associated with frequent tumors (i.e. breast, prostate, colorectal, gastric, and CLL) in Spain[107]. Between 2010 and 2013, CLL cases aged 20 to 85 years were recruited in 11 Spanish hospitals from 5 Spanish provinces (Asturias, Barcelona, Cantabria, Girona and Granada) (**Figure 7**).

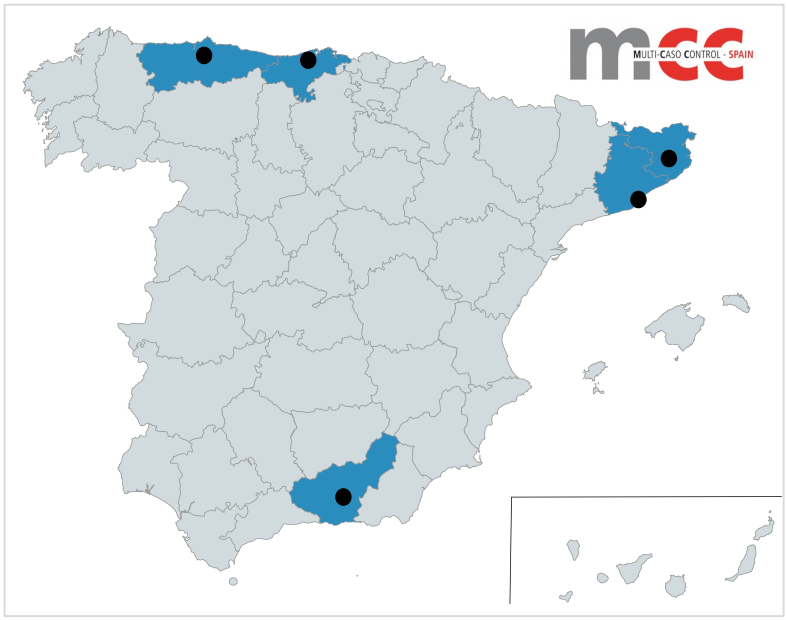


Figure 7: Provinces (in blue) involved in the collection of CLL cases in the MCC-Spain study.

Simultaneously, population-based controls frequency-matched to cases, by age (in 5-year intervals), sex and province of recruitment were randomly selected from primary care centers within the hospitals' catchment areas. After applying specific study-specific exclusion criteria, a total of 1,605 controls and 369 CLL cases were included in the current analysis. All participants signed an informed consent, and approval for the study was obtained from the ethical review boards of all recruiting centers.

4.2.2 Outcome definition

CLL cases were diagnosed according the 'International Workshop on CLL criteria'[108] and morphologically and immunologically confirmed using flow cytometry immunophenotype and complete blood cells count. CLL and SLL were considered the same underlying disease[5]. Given the indolent course of the disease, CLL cases were recruited and interviewed within 3 years from diagnosis. Disease severity at interview was evaluated using the Rai staging system obtained from medical records and verified by local hematologists.

4.2.3 Data collection

Data on socio-demographic factors, lifestyle and personal/family medical history were collected through face-to-face interviews performed by trained personnel. Height and weight at different ages were self-reported. The questionnaire in Spanish is available at www.mccspain.org.

In addition, subjects were provided a semi-quantitative FFQ, which was a modified version from a previously validated instrument in Spain to include regional products[109]. The FFQ was self-administered and returned by mail or filled out face to face. It included 140 food items with portion sizes specified for each item, and assessed usual dietary intake during the previous year. Cross-check questions on aggregated food group consumption were used to adjust the frequency of food consumption and reduce misreporting of food groups with large numbers of items[110]. Nutrient intakes were estimated using food composition tables published for Spain, and other sources[111].

4.2.4 Exposure assessment

Three dietary patterns identified in the Spanish case-control study EpiGEICAM[99] were reconstructed in the MCC-study:

- **Western dietary pattern** characterized by high intakes of high-fat dairy products, processed meat, refined grains, sweets, caloric drinks, convenience food and sauces.
- **Prudent pattern**, with high intakes of low-fat dairy products, vegetables, fruits, whole grains and juices.
- **Mediterranean-like pattern**, defined by a high intake of fish, vegetables, legumes, boiled potatoes, fruits, olives and vegetable oil.

Dietary information extracted from a semi-quantitative FFQ in the EpiGEICAM study was converted to mean daily intake in grams and grouped into 26 food categories (**Table 9**). Major existing dietary patterns were identified in the control population by applying PCA over the 26 inter-correlated food groups. The obtained set of loadings (**Figure 8**) represent the correlation between the consumption of each food group and the component/pattern score and can be used to apply such patterns in other populations[112].

In the MCC-study, we grouped the FFQ items into the same 26 food groups, and calculated the score of adherence to the Western, Prudent and Mediterranean dietary patterns as a linear combination of the loads described in the EpiGEICAM study and the log-transformed centered food group consumption reported by the participants of MCC-Spain study.

Table 9. Composition of the 26 food groups and component loading for each pattern identified in the EpiGEICAM study (adapted from Castelló *et al.* 2014[99]).

Food group	It includes:	L_W[1]	L_P[1]	L_M[1]
High-fat dairy	Whole-fat milk, condensed milk, whole-fat yogurt, semi-cured, cured, or creamy cheese, blue cheese, custard, milk shake, ice-cream, double cream.	0.60	−0.11	0.20
Low-fat dairy	Semi-skimmed and skimmed milk, soy milk, skimmed yogurt, curd, cottage or fresh white cheese.	−0.49	0.60	−0.01
Eggs	Eggs.	0.19	0.08	0.16
White meat	Chicken, rabbit and duck.	0.08	0.17	0.18
Red meat	Pork, beef, lamb, liver (beef, pork or chicken), entrails, hamburgers (pork or beef) and meatballs (pork or beef).	0.27	0.09	0.22
Processed meat	Sausages, serrano ham and other cold meat, bacon, pâté, foie-gras.	0.36	0.10	0.26
White fish	Fresh or frozen white fish (hake, sea bass, sea bream), ½·salted fish and ½·smoked fish.	0.01	0.24	0.34
Oily fish	Fresh or frozen blue fish (tuna, swordfish, sardines, anchovies, salmon), canned fish, ½·salted fish and ½·smoked fish.	0.05	0.24	0.44
Seafood/shellfish	Clams, mussels, oysters, squid, cuttlefish, octopus, prawn, crab, shrimp and similar products.	0.17	0.27	0.35
Leafy vegetables	Spinach, chard, lettuce and other leafy vegetables.	−0.11	0.34	0.40
Fruiting vegetables	Tomato, eggplant, zucchini, cucumber, pepper, artichoke and avocado.	0.00	0.36	0.45
Root vegetables	Carrot, pumpkin and radish.	0.05	0.35	0.44
Other vegetables	Cooked cabbage, cauliflower or broccoli, onion, green beans, asparagus, mushrooms, corn, garlic, gazpacho, vegetable soup and other vegetables.	−0.04	0.40	0.42
Legumes	Peas, lentils, chickpeas, beans and broad beans.	0.21	0.15	0.34
Potatoes	Roasted or boiled potatoes and sweet potatoes.	0.17	0.25	0.40
Fruits	Orange, grapefruit, mandarin, banana, apple, pear, grapes, kiwi, strawberries, cherries, peach, figs, melon or watermelon, prunes, mango and papaya and other fresh or dried fruits.	−0.07	0.31	0.31
Nuts	Almonds, peanuts, pine nuts, hazelnut	0.18	0.22	0.29
Refined grains	White-flour bread, rice, pasta	0.37	0.15	0.23
Whole grains	Whole-grain bread and breakfast cereals	−0.43	0.47	−0.06
Olives and vegetable oil	Olives, added olive oil to salads, bread and dishes, other vegetable oils (sunflower, corn, and	0.12	0.19	0.34

	soybean).			
Other edible fats	Margarine, butter and lard.	0.22	0.02	0.11
Sweets	Chocolate and other sweets, cocoa powder, plain cookies, chocolate cookies, pastries (croissant, donut, cake, pie or similar)	0.35	0.18	0.05
Sugary	Jam, honey, sugar and fruit in sugar syrup.	0.24	0.05	0.00
Juices	Tomato juice, freshly squeezed orange juice, juice (other than freshly squeezed)	0.25	0.67	−0.39
Caloric drinks	Sugar-sweetened soft drinks and nut milk.	0.74	0.21	−0.25
Convenience food and sauces	Croquette, fish sticks, dumplings, kebab, fried potatoes, crisps, pizza, instant soup, mayonnaise, tomato sauce, hot sauce, ketchup and other sauces.	0.47	0.12	0.24

[1]Component loadings for the W: Western; P: Prudent; M: Mediterranean dietary patterns.

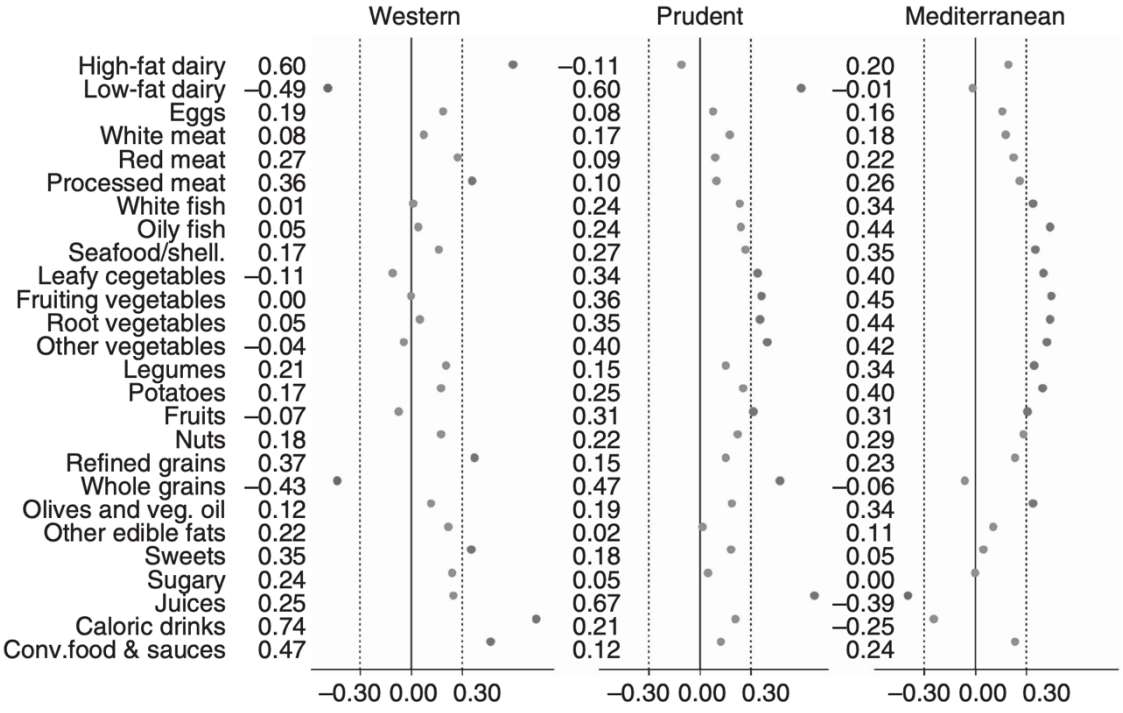


Figure 8. Western, Prudent, and Mediterranean dietary patterns extracted from the EPIGEICAM study (adapted from Castello *et al.* 2014[99]). Each pattern, derived by principal component analysis (PCA), is defined by 26-food group loads. Adherence to each pattern is obtained by multiplying the intake of each food group by its respective load, which are finally summed into a linear score.

4.2.5 Statistical analyses

The association between the dietary patterns and CLL was evaluated using mixed logistic regression models with random province-specific intercepts. The three dietary patterns were included in the model both as continuous variables (1-SD increase of the controls' scores) and as categorical variables (according to the quartile distribution in all controls). All models were adjusted for age (years, continuous), sex, education (no formal education, primary school, secondary school, university), and energy intake (kcal/day, continuous), and province of residence was included as a random effect term.

A possible effect modification of sex, BMI, energy intake, tobacco, physical activity, working on a farm and family history of hematologic malignancies was tested using log-likelihood ratio tests, including an interaction term between each pattern (as continuous) and such variables. In addition, results by Rai stage (0 vs. I-IV) were calculated with multinomial logistic regression models using the same adjustment.

Finally, we performed sensitivity analyses to examine if inclusion of cases with longer period of time from diagnosis to recruitment (≥ 1 year) might have caused potential reverse causality. In addition, we checked if cases treated before the interview might have also affected the overall estimates.

The p for heterogeneity of effects across Rai stage and for sensitivity analyses was obtained with the Wald test. All analyses were performed using STATA/MP (version 14.1, 2015, StataCorp LP) and statistical significance was set at 2-sided $p<0.05$.

5. RESULTS

The results of this thesis are presented as three original articles:

Paper I: Solans M, Benavente Y, Saez M, *et al.* Adherence to the Mediterranean diet and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2018 Dec 26. doi: 10.1002/ijc.32091. [Epub ahead of print]
[**2017 Impact factor: 7.360**; Q1 Oncology, position 23 of 223].

Paper II: Solans M, Benavente Y, Saez M, *et al.* Inflammatory potential of diet and risk of lymphoma in the European Prospective Investigation into Cancer and Nutrition. Eur J Nutr. 2019 Mar 22. doi: 10.1007/s00394-019-01947-0. [Epub ahead of print]
[**2017 Impact factor: 4.423**; Q1 Nutrition and Dietetics, position 14 of 83].

Paper III: Solans M, Castelló A, Benavente Y, *et al.* Adherence to the Western, Prudent, and Mediterranean dietary patterns and chronic lymphocytic leukemia in the MCC-Spain study. Haematologica. 2018 103:1881-1888.
[**2017 Impact factor: 9.090;** Q1 Hematology, position 5 of 71].

5.1Paper I

Adherence to the Mediterranean diet and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition

Solans M, Benavente Y, Saez M, Agudo A, Naudin S, Saberi Hosnijeh F, Noh H, Freisling H, Ferrari P, Besson C, Mahamat-Saleh Y, Boutron-Ruault MC, Kühn T, Kaaks R, Boeing H, Lasheras C, Rodríguez-Barranco M, Amiano P, Huerta JM, Barricarte A, Schmidt J, Vineis P, Riboli E, Trichopoulou A, Bamia C, Peppa E, Masala G, Agnoli C, Tumino R, Sacerdote C, Panico S, Skeie G, Weiderpass E, Jerkeman M, Ericson U, Späth F, Nilsson LM, Dahm CC, Overvad K, Bolvig AK, Tjønneland A, de Sanjose S, Buckland G, Vermeulen, Nieters A, Casabonne D.

Int J Cancer. 2018 Dec 26. doi: 10.1002/ijc.32091. [Epub ahead of print]

Box 2 | Overview of paper I

What is already known on this subject

- There is growing evidence of the protective role of the Mediterranean diet on solid neoplasms.
- The association between adherence to a Mediterranean diet and risk of lymphoma remains unexplored.

What this study adds

- For the first time, a prospective study has shown that high adherence to a Mediterranean diet is associated with a 2% lower risk of lymphoma for each 1-SD increase in the arMED score. No specific-subtype associations were found.
- Whilst its modest effect, the Mediterranean dietary pattern appears to be valuable for primary prevention of lymphoma in healthy populations.

Solans M, Benavente Y, Saez M, Agudo A, Naudin S, Saberi Hosnijeh F, Noh H, Freisling H, et al. "Adherence to the Mediterranean diet and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition". *International Journal of Cancer.* Vol 145, Issue 1 (2019) : 122-131.

https://doi.org/10.1002/ijc.32091

First published:26 December 2018

Abstract

There is a growing evidence of the protective role of the Mediterranean diet (MD) on cancer. However, no prospective study has yet investigated its influence on lymphoma. We evaluated the association between adherence to the MD and risk of lymphoma and its subtypes in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. The analysis included 476,160 participants, recruited from 10 European countries between 1991 and 2001. Adherence to the MD was estimated through the adapted relative MD (arMED) score excluding alcohol. Cox proportional hazards regression models were used while adjusting for potential confounders. During an average follow-up of 13.9 years, 3,136 lymphomas (135 Hodgkin lymphoma [HL], 2,606 non-HL and 395 lymphoma not otherwise specified) were identified. Overall, a 1-unit increase in the arMED score was associated with a 2% lower risk of lymphoma (95% CI: 0.97; 1.00, p-trend = 0.03) while a statistically nonsignificant inverse association between a high *versus* low arMED score and risk of lymphoma was observed (hazard ratio [HR]: 0.91 [95% CI 0.80; 1.03], p-trend = 0.12). Analyses by lymphoma subtype did not reveal any statistically significant associations. Albeit with small numbers of cases (*N* = 135), a suggestive inverse association was found for HL (HR 1-unit increase = 0.93 [95% CI: 0.86; 1.01], p-trend = 0.07). However, the study may have lacked statistical power to detect small effect sizes for lymphoma subtype. Our findings suggest that an increasing arMED score was inversely related to the risk of overall lymphoma in EPIC but not by subtypes. Further large prospective studies are warranted to confirm these findings.

Keywords

Lymphoma; Mediterranean diet; Europe; prospective studies; risk

5.2Paper II

Inflammatory potential of diet and risk of lymphoma in the European Prospective Investigation into Cancer and Nutrition

Solans M, Benavente Y, Saez M, Agudo A, Jakszyn P, Naudin S, Saberi Hosnijeh F, Gunter M, Huybrechts I, Ferrari P, Besson C, Mahamat-Saleh Y, Boutron-Ruault MC, Kühn T, Kaaks R, Boeing H, Lasheras C, Sánchez MJ, Amiano P, Chirlaque MD, Ardanaz E, Schmidt JA, Vineis PA, Riboli E, Trichopoulou A, Karakatsani A, Valanou E, Masala G, Agnoli C, Tumino R, Sacerdote C, Mattiello A, Skeie G, Weiderpass E, Jerkeman M, Dias JA, Späth F, Nilsson LM, Dahm CC, Overvad K, Petersen KEN, Tjønneland A, de Sanjose S, Vermeulen R, Nieters A, Casabonne D.

Eur J Nutr. 2019 Mar 22. doi: 10.1007/s00394-019-01947-0. [Epub ahead of print].

Box 3 | Overview of paper II

What is already known on this subject

- Chronic inflammation plays a critical role in lymphomagenesis and several dietary factors seem to be involved its regulation.
- Numerous epidemiological studies have found associations between a pro-inflammatory diet, measured through the DII or ISD, and several solid neoplasms.
- To date, however, evidence of the inflammatory potential of diet and risk of hematological malignancies is scarce, with no prospective data available.

What this study adds

- Our findings showed that a high ISD, reflecting a pro-inflammatory diet, is modestly associated with the risk of mature B-cell NHL [HR for a 1-SD increase: 1.07 (95%CI: 1.01; 1.14)].
- Further research involving biomarkers of inflammation would help to better understand the potential underlying mechanisms between diet-related inflammation and lymphomagenesis.

Solans M, Benavente Y, Saez M, Agudo A, Jakszyn P, Naudin S, Saberi Hosnijeh F, et al. “Inflammatory potential of diet and risk of lymphoma in the European Prospective Investigation into Cancer and Nutrition”. *European Journal of Nutrition.* Published online: 22 March 2019

https://doi.org/10.1007/s00394-019-01947-0

Received: 17 October 2018 / Accepted: 11 March 2019

Abstract

Introduction

Chronic inflammation plays a critical role in lymphomagenesis and several dietary factors seem to be involved its regulation. The aim of the current study was to assess the association between the inflammatory potential of the diet and the risk of lymphoma and its subtypes in the European Investigation into Cancer and Nutrition (EPIC) study.

Methods

The analysis included 476,160 subjects with an average follow-up of 13.9 years, during which 3,136 lymphomas (135 Hodgkin lymphoma (HL), 2606 non-Hodgkin lymphoma (NHL) and 395 NOS) were identified. The dietary inflammatory potential was assessed by means of an inflammatory score of the diet (ISD), calculated using 28 dietary components and their corresponding inflammatory weights. The association between the ISD and lymphoma risk was estimated by hazard ratios (HR) and 95% confidence intervals (CI) calculated by multivariable Cox regression models adjusted for potential confounders.

Results

The ISD was not associated with overall lymphoma risk. Among lymphoma subtypes, a positive association between the ISD and mature B-cell NHL (HR for a 1-SD increase: 1.07 (95% CI 1.01; 1.14), *p* trend = 0.03) was observed. No statistically significant association was found among other subtypes. However, albeit with smaller number of cases, a suggestive association was observed for HL (HR for a 1-SD increase = 1.22 (95% CI 0.94; 1.57), *p* trend 0.13).

Conclusions

Our findings suggested that a high ISD score, reflecting a pro-inflammatory diet, was modestly positively associated with the risk of B-cell lymphoma subtypes. Further large prospective studies on low-grade inflammation induced by diet are warranted to confirm these findings.

Keywords

Chronic inflammation · Inflammatory score of the diet · Lymphoma · Nutrition · Prospective studies

5.3Paper III

Adherence to the Western, Prudent, and Mediterranean dietary patterns and chronic lymphocytic leukemia in the MCC-Spain study

Solans M, Castelló A, Benavente Y, Marcos-Gragera R, Amiano P, Gracia-Lavedan E, Costas L, Robles C, Gonzalez-Barca E, de la Banda E, Alonso E, Aymerich M, Campo E, Dierssen-Sotos T, Fernández-Tardón G, Olmedo-Requena R, Gimeno E, Castaño-Vinyals G, Aragonés N, Kogevinas M, de Sanjose S, Pollán M, Casabonne D.

*Haematologica.*2018; 103: 1881-1888; **Doi**:10.3324/haematol.2018.192526

Box 4 | Overview of paper III

What is already known on this subject

- Results from epidemiological research on the effect of individual foods or nutrients on CLL are inconsistent.
- Focusing on dietary patterns could provide additional information for clarifying the role of nutrition in CLL etiology.
- Few studies on overall diet have been conducted in NHL, most of them underpowered to find CLL-specific associations.

What this study adds

- This study provides the first evidence for an association between a Western dietary pattern and CLL [OR per 1-SD increase =1.19 (95% CI: 1.03; 1.37)], independently of RAI stage at diagnosis.
- No associations were found for a Mediterranean-like or Prudent dietary patterns.
- These results suggest that a proportion of CLL cases could be prevented by modifying dietary habits.

ARTICLE

Adherence to the Western, Prudent, and Mediterranean dietary patterns and chronic lymphocytic leukemia in the MCC-Spain study

Marta Solans,[1,2,3*] Adela Castelló,[1,4,5*] Yolanda Benavente,[1,6] Rafael Marcos-Gragera,[2,3] Pilar Amiano,[1,7] Esther Gracia-Lavedan,[1,8,9] Laura Costas,[6] Claudia Robles,[10] Eva Gonzalez-Barca,[11] Esmeralda de la Banda,[12] Esther Alonso,[12] Marta Aymerich,[13] Elias Campo,[13] Trinidad Dierssen-Sotos,[1,14] Guillermo Fernández-Tardón,[1,15] Rocio Olmedo-Requena,[1,16,17] Eva Gimeno,[18] Gemma Castaño-Vinyals,[1,8,9,19,] Nuria Aragonés,[1,20] Manolis Kogevinas,[1,8,9,19] Silvia de Sanjose,[1,6,21] Marina Pollán[1,4] and Delphine Casabonne[1,6**]**

[1]Centro de Investigación Biomédica en Red: Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; [2]Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Spain; [3]Epidemiology Unit and Girona Cancer Registry, Catalan Institute of Oncology, Girona, Spain; [4]Cancer Epidemiology Unit, National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain; [5]Faculty of Medicine, University of Alcalá, Alcalá de Henares, Madrid, Spain; [6]Unit of Molecular Epidemiology and Genetic in Infections and Cancer (UNIC-Molecular), Cancer Epidemiology Research Programme (IDIBELL), Catalan Institute of Oncology, L' Hospitalet De Llobregat, Spain; [7]Public Health Division of Gipuzkoa, BioDonostia Research Institute, San Sebastian, Spain; [8]ISGlobal, Barcelona, Spain; [9]Universitat Pompeu Fabra (UPF), Barcelona, Spain; [10]Unit of Information and Interventions in Infections and Cancer (UNIC-I&I), Cancer Epidemiology Research Programme, (IDIBELL), Catalan Institute of Oncology, L' Hospitalet De Llobregat, Spain; [11]Hematology, Bellvitge Biomedical Research Institute (IDIBELL), Catalan Institute of Oncology, L'Hospitalet de Llobregat, Spain; [12]Hematology Laboratory, Department of Pathology, Hospital Universitari de Bellvitge, L'Hospitalet de Llobregat, Spain; [13]Hospital Clinic de Barcelona, University of Barcelona, CIBERONC, Barcelona Spain; [14]University of Cantabria - Marqués de Valdecilla Research Institute (IDIVAL), Santander, Spain; [15]University Institute of Oncology (IUOPA), University of Oviedo, Spain; [16]Department of Preventive Medicine and Public Health, University of Granada, Spain; [17]Instituto de Investigación Biosanitaria de Granada, Hospitales Universitarios de Granada, Spain; [18]Hematology Department, Hospital del Mar, Barcelona, Spain; [19]Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain; [20]Epidemiology Section, Public Health Division, Department of Health of Madrid, Spain and [21]PATH, Reproductive Health, Seattle, WA, USA

**MS and AC contributed equally to this work. **MP and DC contributed equally to this work.*

Haematologica 2018
Volume 103(11):1881-1888

Correspondence:
dcasabonne@iconcologia.net

Received: March 6, 2018.
Accepted: June 25, 2018.
Pre-published: June 28, 2018.

doi:10.3324/haematol.2018.192526

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/103/11/1881

ABSTRACT

Diet is a modifiable risk factor for several neoplasms but evidence for chronic lymphocytic leukemia (CLL) is sparse. Previous studies examining the association between single-food items and CLL risk have yielded mixed results, while few studies have been conducted on overall diet, reporting inconclusive findings. This study aimed to evaluate the association between adherence to three dietary patterns and CLL in the multicase-control study (MCC-Spain) study. Anthropometric, sociodemographic, medical and dietary information was collected for 369 CLL cases and 1605 controls. Three validated dietary patterns, Western, Prudent and Mediterranean, were reconstructed in the MCC-Spain data. The association between adherence to each dietary pattern and CLL was assessed, overall and by Rai stage, using mixed logistic regression models adjusted for potential confounders. High adherence to a Western dietary pattern (i.e. high intake of high-fat dairy products, processed meat, refined grains, sweets, caloric drinks, and convenience food) was associated with CLL [ORQ4 *vs.* Q1=1.63 (95%CI 1.11; 2.39); *P*-trend=0.02; OR 1-SD increase=1.19 (95%CI: 1.03; 1.37)], independently of Rai stages. No differences in the association were observed according to sex, Body Mass Index, energy intake, tobacco, physical activity, working on a farm, or family history of hematologic malignancies. No associations were observed for Mediterranean and Prudent dietary patterns and CLL. This study provides the first evidence for an association between a Western dietary pattern and CLL, suggesting that a proportion of CLL cases could be prevented by modifying dietary habits. Further research, especially with a prospective design, is warranted to confirm these findings.

Introduction

Chronic lymphocytic leukemia (CLL) is the commonest leukemia among the adult population in Western countries, with an annual incidence rate of around 5 per 100,000 person-years in Europe,[1] but its etiology is still poorly understood. A pooled analysis of 2440 CLL cases and 15,186 controls from the InterLymph consortium showed significant inverse associations with atopic conditions, smoking, blood transfusion history, and recreational sun exposure, and positive associations with height, hepatitis C virus seropositivity, living or working on a farm, working as a hairdresser, and family history of hematologic malignancies.[2]

Diet is a modifiable risk factor for several neoplasms,[3] but evidence for CLL is inconclusive. Epidemiological data on the association of diet and CLL are heterogeneous, and mainly arise from studies on nutrients or single food items. While most prospective studies[4-12] did not find any association with a wide range of dietary factors, case-control studies[13-25] have yielded contradictory results for items such as meat, dairy or vegetable intake. Some authors argue that focusing on overall dietary patterns instead of on individual foods or nutrients may better capture dietary variability in the population's diet while allowing the evaluation of interactions between dietary factors.[26] However, the few studies that have been conducted on overall diet and CLL[25,27,28] reported inconclusive findings, mainly due to small sample size.

A population-based multicase-control study (MCC-Spain) was launched to evaluate the influence of environmental exposures and their interaction with genetic factors in CLL, among other cancers.[29] The aim of the present study was to evaluate the association between adherence to three validated dietary patterns,[30] Western, Prudent and Mediterranean, and CLL in the MCC-Spain study.

Methods

Study population

MCC-Spain is a multicentric case-control study with population controls and cases with common tumors (prostate, breast, colorectal, gastroesophageal and CLL) in Spain. Between 2010 and 2013, CLL cases aged 20-85 years were recruited in 11 Spanish hospitals from 5 Spanish provinces (Asturias, Barcelona, Cantabria, Girona and Granada). Simultaneously, population-based controls frequency-matched to cases according to age (5-year intervals), sex, and province of recruitment were randomly selected from primary care centers within the hospitals' catchment areas. Participation rates were 87% in cases and 53% in controls, with variability among geographical regions. After applying specific diet exclusion criteria (excluding participants with no dietary data or with missing or implausible energy intakes under 750 or over 4500 kcal/day), a total of 1605 controls and 369 CLL cases were included in the study. All participants gave informed consent. Approval for the study was obtained from the ethical review boards of all recruiting centers. Additional information regarding the study design has been provided elsewhere.[29]

Outcome definition

Chronic lymphocytic leukemia cases were diagnosed according to the International Workshop on CLL criteria: presence of an absolute count ≥5 x10^9 B cells/L for three or more months in peripheral blood and a clonal population of $CD5^+$, $CD19^+$, and $CD23^+$ B cells.[31] All diagnoses were morphologically and immunologically confirmed using flow cytometry immunophenotype and complete blood cell count. CLL and small lymphocytic lymphoma were considered the same underlying disease. Given the indolent course of the disease, CLL cases were recruited and interviewed within three years from diagnosis. Disease severity at interview was evaluated using the Rai staging system obtained from medical records and verified by local hematologists. CLL subjects were then categorized into two groups based on Rai stage: a) low-risk category including asymptomatic patients with lymphocytosis only (Rai 0); and b) intermediate/high-risk category including patients with either lymphadenopathy, hepatomegaly, splenomegaly, anemia and/or thrombocytopenia (Rai I-IV).

Data collection

Data on socio-demographic factors, lifestyle and personal/family medical history were collected through face-to-face interviews performed by trained personnel. Height and weight at different ages were self-reported. The questionnaire in Spanish is available at *www.mccspain.org*.

In addition, subjects were provided a semi-quantitative Food Frequency Questionnaire (FFQ), which was a modified version from a previous tool validated in Spain to include regional products.[32] The FFQ was self-administered and returned by mail or filled out face-to-face. It included 140 food items with portion sizes specified for each item, and assessed usual dietary intake during the previous year. Cross-check questions on aggregated food group consumption were used to adjust the frequency of food consumption and reduce misreporting of food groups with large numbers of items.[33] Nutrient intakes were estimated using food composition tables published for Spain, and other sources.[34] The response rate of the FFQ was slightly lower in cases (82%) than in controls (87%). Overall, responsiveness was not associated with age, and individuals from Granada were less likely to answer the FFQ than those from Barcelona. Those individuals who did not answer the diet questionnaire had a lower level of education and, in controls, were also more likely to be women.

Dietary patterns

Three validated dietary patterns identified in a Spanish case-control study (EpiGEICAM)[30] were reconstructed in the MCC-study: a) a Western dietary pattern characterized by high intake of high-fat dairy products, processed meat, refined grains, sweets, caloric drinks, convenience food and sauces; b) a Prudent pattern, with high intake of low-fat dairy products, vegetables, fruits, whole grains and juices; and c) a Mediterranean pattern, defined by a high intake of fish, vegetables, legumes, boiled potatoes, fruits, olives, and vegetable oil. Further information on identification of the dietary patterns can be found elsewhere.[30] In brief, dietary information extracted from a semi-quantitative questionnaire in the EpiGEICAM study was converted to mean daily intake in grams and grouped into 26 food categories. Major existing dietary patterns were identified in the control population by applying principal components analysis (PCA) without rotation of the variance-covariance matrix over the 26 inter-correlated food groups. The set of loadings obtained represent the correlation between the consumption of each food group and the component/pattern score, and can be used to apply such patterns to other populations.[35] In the MCC-study, we grouped the FFQ items into the same 26 food groups (*Online Supplementary Table S1*), and calculated the score of adherence to the Western, Prudent and Mediterranean dietary patterns as a linear combination of the loads described in the EpiGEICAM study and the log-transformed centered food group consumption reported by the participants of MCC-Spain study.

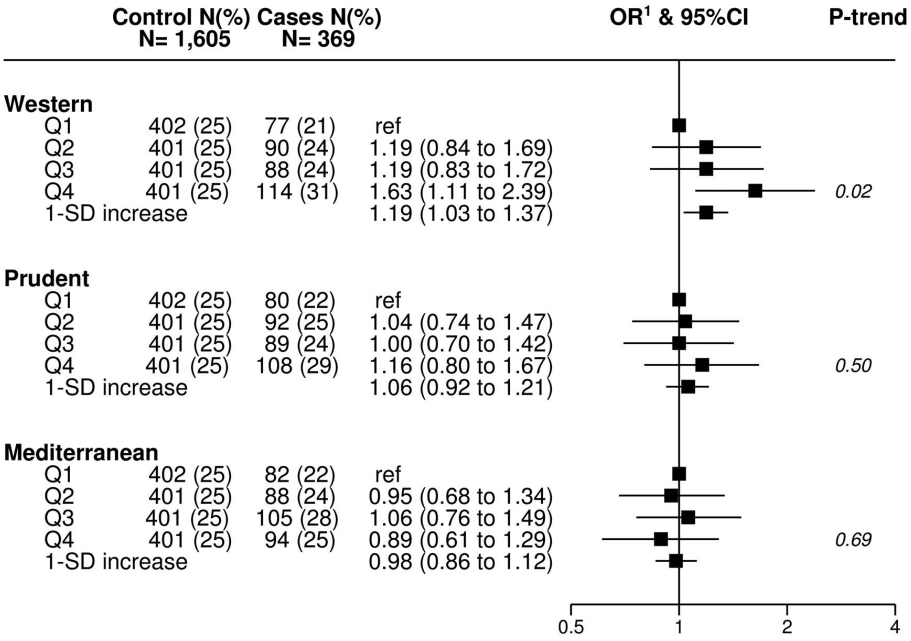


Figure 1. Association between adherence to dietary patterns and chronic lymphocytic leukemia in the multicase-control (MCC-Spain) study. OR: Odds Ratio; 95%CI: 95% confidence interval; Q: quartile; SD: standard deviation. Black squares indicate OR and horizontal lines represent 95%CI. [1]Logistic regression models adjusted for age, sex, education, energy intake (kcal/day) with province of residence as random effect.

Statistical analysis

As descriptive analyses, we compared anthropometric, sociodemographic and lifestyle characteristics between cases and controls. χ^2 test was used to evaluate the level of significance of the differences observed in categorical variables, Student *t*-test for normally distributed continuous variables, and Wilcoxon rank-sum test for non-normally distributed continuous variables. In addition, we analyzed the distribution of each dietary pattern (continuous) across categories of descriptive variables. Student *t*-test was used to assess differences observed in variables with two categories and ANOVA for those with more than two categories.

The association between the dietary patterns and CLL was evaluated using mixed logistic regression models with random province-specific intercepts. The exposure variables (adherence to Western, Prudent or Mediterranean patterns) were included in the model both as continuous variables [1-standard deviation (SD) increase in the controls' scores] and as categorical variables (according to the quartile distribution in all controls). All models were adjusted for age (years, continuous), sex, education (no formal education, primary school, secondary school, university), and energy intake (kcal/day, continuous) as fixed effects and province of residence as a random effect term. Height (cm, continuous), waist-to-hip ratio (continuous), Body Mass Index (BMI in kg/m^2, continuous), experience working on a farm (yes, no), family history of hematologic malignancies (yes, no), alcohol consumption (g/day, continuous), smoking status (never, past, current), and physical activity [in the last 10 years, measured in Metabolic Equivalent of Task (METs)/week: inactive (0), low (0.1-8), moderate (8-15.9) and very active (≥16)] were examined as potential confounders, but were not included in the final models as they were not found alone, or in combination, to affect the estimates. Interaction terms were modeled between each of these separate variables and the dietary score (continuous), and tested using log-likelihood ratio tests. A possible effect modification of sex, BMI, energy intake, tobacco, physical activity, working on a farm, and family history of hematologic malignancies was tested including an interaction term between each of the patterns and such variables. The estimation of the effects according to Rai stage (0 *vs.* I-IV) was calculated with multinomial logistic regression models adjusted by the set of variables described above plus province of residence as random effect term. Finally, sensitivity analyses were performed to examine how the inclusion of: i) cases with longer period of time from diagnosis to recruitment (<1 year *vs.* ≥ 1 year); and ii) cases treated before the interview affected the overall estimates. Odds Ratios (OR) and 95% confidence intervals (CI) were also obtained with multinomial logistic regression models. The *P*-value for heterogeneity of effects across Rai stage and for sensitivity analyses was obtained with the Wald test. All analyses were performed using STATA/MP (v.14.1, 2015, StataCorp LP) and statistical significance was set at two-sided $P<0.05$.

Results

Distribution of baseline characteristics between cases and controls is shown in Table 1. Compared with controls, cases were more adherent to the Western pattern, while no differences in level of adherence to the Prudent and Mediterranean patterns were observed in bivariate analyses. CLL cases were also slightly older, had a higher waist-to-hip ratio, and were more likely to have a family history of hematologic malignancy and to have worked on a farm. No other differences were observed for any of the other pre-selected variables.

The distribution of key characteristics of controls according to level of adherence to each dietary pattern is shown in *Online Supplementary Figure S1*. Controls with greater adherence to a Western pattern were more likely to be men, younger, taller, current smokers, less prone to have worked in farming or agriculture, had a lower BMI and waist-to-hip ratio, and a higher level of education, energy and alcohol intake. Those with a higher adherence to a Prudent pattern were more likely to be women, younger, taller, physically active, never/former smokers,

Table 1. Baseline characteristics of chronic lymphocytic leukemia for cases and controls in the multicase-control (MCC-Spain) study.

	Controls (n=1605)	Cases (n=369)	P[1]
Western, mean (SD)	5.88 (1.46)	6.06 (1.40)	0.03
Prudent, mean (SD)	6.55 (1.12)	6.66 (1.03)	0.07
Mediterranean, mean (SD)	7.08 (1.00)	7.16 (0.88)	0.20
Province, n(%)			**<0.001**
Barcelona	900 (56)	242 (66)	
Asturias	211 (13)	51 (14)	
Cantabria	281 (18)	21 (6)	
Granada	144 (9)	27 (7)	
Girona	69 (4)	28 (8)	
Age (years), mean (SD)	64.30 (10.54)	66.19 (10.12)	**0.002**
Sex, n(%)			0.95
Male	936 (58)	217 (59)	
Female	669 (42)	152 (41)	
Energy intake (kcal/day), mean (SD)	1901.15 (585.88)	1937.91 (612.06)	0.28
Current alcohol intake (g/day), median (IQI)[2]	8.80 (0.58;27.32)	8.82 (0.83;24.47)	0.75
BMI (kg/m^2), mean (SD)[2]	26.99 (4.50)	27.32 (4.43)	0.21
Height (cm), mean(SD)[2]	165.63 (8.51)	165.97 (9.11)	0.50
Waist-to-hip ratio[3], n(%)			**0.004**
Low	460 (29)	77 (21)	
Moderate	449 (28)	98 (27)	
High	682 (42)	192 (52)	
Unknown	14 (1)	2 (1)	
Smoking status, n(%)			0.39
Never	696 (43)	165 (45)	
Former	602 (38)	134 (36)	
Current	303 (19)	67 (18)	
Unknown	4 (0)	3 (1)	
Education, n(%)			0.54
No formal education	357 (22)	94 (25)	
Primary	502 (31)	106 (29)	
Secondary	461 (29)	107 (29)	
University	285 (18)	62 (17)	
Physical activity[4], n(%)			0.50
Inactive	656 (41)	136 (37)	
Low	219 (14)	55 (15)	
Moderate	190 (12)	47 (13)	
Very active	502 (31)	118 (32)	
Unknown	38 (2)	13 (4)	
Ever worked in farming or agriculture, n(%)			**<0.001**
No	1257 (78)	258 (70)	
Yes	323 (20)	108 (29)	
Unknown	25 (2)	3 (1)	
Family history of hematologic malignancy, n(%)			**<0.001**
No	1551 (97)	333 (90)	
Yes	54 (3)	36 (10)	
Rai stage			
0	–	199 (54)	
I-IV	–	150 (41)	
Unknown	–	20 (5)	

SD: standard deviation; IQI: interquartile interval; BMI: Body Mass Index. [1]*P*-value for heterogeneity calculated with the Student *t*-test for comparison of normally distributed continuous variables, with the Wilcoxon rank-sum test for comparison of non-normally distributed continuous variables (alcohol intake), and with the χ^2 test for categorical variables. [2]% of missing values in continuous variables: alcohol intake (2%), BMI (4%), height (3%). [3]Waist-to-hip ratio risk categories according to WHO criteria. [4]Physical activity, in the last ten years, measured in METs/week: inactive (0), low (0.1-8), moderate (8-15.9), and very active (≥16). In bold: $P<0.05$.

Table 2. Association between adherence to dietary patterns and chronic lymphocytic leukemia by severity of the disease, in the multicase-control (MCC-Spain) study.

		Rai 0		Rai I-IV		
	Controls N(%) (n=1605)	Cases N(%) (n=199)	OR[1] (95% CI)	Cases (n=150)	OR[1] (95% CI)	*P*-het[2]
Western						
Q1	402 (25)	45 (23)	1	26 (17)	1	
Q2	401 (25)	47 (24)	1.11 (0.71;1.74)	39 (26)	1.40 (0.82;2.38)	
Q3	401 (25)	47 (24)	1.17 (0.73;1.87)	37 (25)	1.32 (0.76;2.30)	
Q4	401 (25)	60 (30)	1.60 (0.97;2.65)	48 (32)	1.71 (0.95;3.06)	
P-trend			0.07		0.11	
1-SD increase			1.15 (0.95;1.39)		1.26 (1.02;1.56)	0.50
Prudent						
Q1	402 (25)	42 (21)	1	30 (20)	1	
Q2	401 (25)	53 (27)	1.09 (0.70;1.70)	37 (25)	1.16 (0.69;1.93)	
Q3	401 (25)	52 (26)	1.07 (0.68;1.69)	32 (21)	0.97 (0.56;1.66)	
Q4	401 (25)	52 (26)	1.01 (0.62;1.64)	51 (34)	1.47 (0.86;2.51)	
P-trend			0.98		0.23	
1-SD increase			0.99 (0.83;1.18)		1.18 (0.96;1.45)	0.17
Mediterranean						
Q1	402 (25)	48 (24)	1	29 (19)	1	
Q2	401 (25)	51 (26)	0.88 (0.57;1.36)	29 (19)	0.94 (0.55;1.62)	
Q3	401 (25)	55 (28)	0.88 (0.57;1.37)	46 (31)	1.39 (0.83;2.30)	
Q4	401 (25)	45 (23)	0.65 (0.40;1.06)	46 (31)	1.32 (0.76;2.27)	
P-trend			0.11		0.17	
1-SD increase			0.88 (0.74;1.04)		1.15 (0.93;1.41)	**0.04**

OR: Odds Ratio; 95% CI: 95% Confidence Interval; Q: quartile; SD: Standard Deviation. [1]Multinomial logistical regression models adjusted for age, sex, education, energy intake (kcal/day) with province of residence as a random effect. [2]*P*-value for the heterogeneity of effects. In bold: *P*<0.05.

more highly educated, less prone to have worked in farming or agriculture, and with a higher energy intake and lower alcohol consumption. Finally, controls with a greater adherence to a Mediterranean pattern were more likely to be men, physically active, showing a lower proportion of ever smokers and having worked in farming or agriculture, and a higher energy intake.

Figure 1 summarizes the adjusted ORs for the association between CLL and level of adherence to the Western, Prudent and Mediterranean dietary patterns. Individuals in the highest quartile of the Western score had an OR for CLL of 1.63 (95%CI: 1.11; 2.39) compared with individuals with low adherence (*P* for trend 0.02). Each SD increment in the score was associated with a 19% higher OR of having CLL (95%CI: 1.03; 1.37). No associations were observed for Mediterranean and Prudent diet patterns. The impact of each individual covariate (region, age, sex, education, and energy intake) in the association of the three dietary patterns and CLL is provided in *Online Supplementary Figure S2.*

Since CLL is more prevalent in men, who are also more likely to adhere to a Western dietary pattern (Table 1 and *Online Supplementary Figure S1*), all analyses were stratified according to sex. No differences across sexes were observed for any of the dietary patterns [*P*-heterogeneity (*P*-het): Western (0.79), Prudent (0.11) and Mediterranean (0.17); *data not shown*]. In addition, no differences were observed according to BMI, energy intake, tobacco, physical activity, working on a farm, and family history of hematologic malignancies (all *P* for interaction >0.05; *data not shown*).

Analyses according to Rai-stage did not show significant heterogeneity of effects for the Western or Prudent dietary patterns (*P*-het=0.50 and 0.17, respectively). However, weak opposite trends in relation to a Mediterranean diet pattern were observed; it was inversely associated (although not statistically significant) with Rai 0 CLL [OR 1-SD increase= 0.88 (95%CI: 0.74; 1.04)] and positively related with Rai I-IV CLL [OR 1-SD increase= 1.15 (95%CI: 0.93; 1.41)] (*P*-het=0.04) (Table 2).

Sensitivity analyses according to time from diagnosis to recruitment yielded similar results for the three dietary patterns (*Online Supplementary Table S2*). Similarly, excluding cases treated prior to consent (n=79) did not materially modify the results [*P*-het in trends: Western (0.25), Prudent (0.32) and Mediterranean (0.33)], but higher ORs for a Western dietary pattern were observed in cases treated prior to consent in comparison to those not treated (*Online Supplementary Table S3*).

Discussion

This study provides, for the first time, evidence of an association between adherence to a Western dietary pattern and CLL. By contrast, no associations were found for a Prudent or Mediterranean pattern.

There is limited evidence linking extrinsic-risk factors,

and particularly diet, with non-Hodgkin lymphoma (NHL). In the 2007 report by the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR),[3] the Panel decided not to make any judgements regarding the causality of associations between specific dietary factors and NHL, but pointed out a suggestive inverse association with vegetables, fruit, and alcoholic beverages, and a positive association with meat, total fat, body fatness, and dairy. Recent meta-analyses further support these associations,[36,37] but there is still not sufficient evidence to establish a causal role. Similarly, data on the association of diet and CLL are inconclusive and mainly arise from studies on nutrients or single food items. To our knowledge, 9 prospective studies[4-12] and 13 case-control studies[13-25] have been published on this topic. With the exception of a few studies that found positive associations with consumption of processed meat and poultry,[4] total carbohydrate[8] or fat (in women)[11] intake, and inverse associations with isoflavones consumption,[10] generally large prospective studies found no associations between a wide range of dietary factors and CLL. By contrast, case-control studies have yielded contradictory results for meat,[13,14,16,24,25] dairy products,[13-17] fish[15,18] or vegetables and fruit[14,16,19,22,23,25] intake.

Inconsistencies in previous epidemiological nutritional studies in part reflect the difficulty in disentangling the influence of single food items that, when consumed in combination, may be highly correlated and exert synergistic or antagonistic effects on CLL risk. The examination of dietary patterns, which better reflect the complexity of dietary intake, has been used to address such limitations.[26] So far, only a few studies have examined associations of dietary patterns and risk of CLL,[25,27,28] reporting inconclusive findings, mainly due to small sample size. Ollberding *et al.*[28] pointed out that a high adherence to a 'Meat, Fat and Sweets' dietary pattern, characterized by a high intake of French fries, red meat, processed meat, pizza, salty snacks, sweets and desserts, was associated with an increased risk of overall NHL (ORQ4 *vs.* Q1=3.6; 95%CI: 1.9, 6.8) in a Nebraska case-control study. This association was maintained when stratifying according to lymphoma subtypes but sub-analyses did not include CLL cases due to sample size (n=25). By contrast, a large prospective cohort in the US did not find associations with 'Fat and Meat' pattern and CLL etiology.[27] However, this pattern did not include sweets and deserts, sweetened beverages, or convenience foods, which may be important contributing factors of these associations. Thus, not only differences in the study design and setting, but also in food groups loaded in these data-driven analyses, should be carefully considered when comparing results. In line with our findings, no associations with overall NHL[25,27,28] or CLL[25,27] were detected for a 'healthy' dietary pattern characterized by high intake of fruit and vegetables.

We observed opposite trends in relation to a Mediterranean diet pattern and Rai stages, with stronger adherence among cases with higher disease severity (*P*-het=0.04). We hypothesize that reverse causality could partly explain these results. While Rai 0 patients are usually diagnosed in a routine blood test and present an indolent course, Rai I-IV are more prone to be symptomatic (e.g. night sweat, fatigue, weight loss or fever) and to receive active treatment. Thus, those patients with a more severe disease (and probably more concerned about their illness) would be more prone to shift towards a healthier dietary pattern. However, these results have to be taken with caution since none of the trends showed statistically significant associations.

Chronic lymphocytic leukemia is the most common leukemia in Western countries while its incidence is much lower in Eastern countries, where it accounts for only 1-3% of NHL in most series.[38] While genetic backgrounds may be responsible for some of the differences in the CLL incidence, some studies have suggested that environmental factors also play an important role. A dramatic increase in CLL incidence in Taiwan in recent years was associated with a strong birth-cohort effect, that corresponded to the Westernization of lifestyle in Taiwan since 1960.[39] In addition, a higher incidence of CLL has been reported among US-born Asians compared to foreign-born Asians, pointing out the influence of environmental factors that change with immigration and acculturation to a Westernized lifestyle.[40] Our results further support the view that adopting a Western diet could partly explain these incidence patterns.

A Western diet has been associated with obesity phenotypes,[41] including a higher waist-to-hip ratio, which has in turn been recently linked to higher OR of CLL, particularly in women, in the MCC-Spain study.[42] Despite the fact that CLL cases showed a higher waist-to-hip ratio than controls in our study, waist-to-hip ratio and BMI were not included as covariates in the final multiple-adjusted model since they did not change risk estimates. Hence, an independent effect of the Western dietary pattern may be contributing to CLL lymphomagenesis, which seems plausible from a mechanistic point of view. On one hand, it has been well-established that dietary changes, and particularly switching from a low-fat, plant polysaccharide-rich diet to a high-fat, high-sugar Western diet, can induce alterations in microbiota composition.[43] Beyond its role in the biosynthesis of key components (e.g. vitamins, essential amino acids or short chain fatty acid byproducts), several studies using germ-free mice suggest that microbiota also plays a fundamental role on the induction, training, and function of the host immune system.[44] Exposure to a Western diet may have selected for a microbiota that lack the resilience and diversity required to establish balanced immune responses, and this phenomenon is proposed to account for some of the dramatic rise in autoimmune and chronic inflammatory disorders found in high-income countries. On the other hand, a diet high in fat, refined grains, red and processed meats, and sweets has been largely associated with higher levels of inflammatory markers[45] and with inflammation-related chronic diseases.[46] In particular in CLL, the strong production of inflammatory cytokines and chemokines accompanied by activation of intra-cellular pro-inflammatory pathways, and the presence of somatic mutations that activate pro-inflammatory signaling pathways, suggest that chronic inflammation plays a pathophysiological role in this disease.[47] Thus, an inflammation-related mechanism may in part underlie the observed associations with CLL, although no research on the inflammatory potential of diet and CLL risk has yet been conducted.

The dietary patterns used in this study were identified using the control population of a multicentric case-control study on female breast cancer in Spain.[30] By contrast, the MCC-Spain study included male participants, who may have different dietary habits. However, this difference does not preclude the application of the original scoring

system over the current sample. Scores of adherence to dietary patterns can be calculated following the exact same rules over different populations, resulting in different levels of adherence while still being valid, as has been recently proved.[35] As a matter of fact, the current dietary patterns had previously been constructed in the MCC-study and a Western dietary pattern was positively associated with gastric,[48] breast[49] and colorectal[50] cancers.

One of the main limitations is the study design since case-control studies are prone to selection and recall biases. Measurement errors in the estimation of food intake due to the use of self-reported FFQ are also of some concern. However, the FFQ was validated in the Spanish population and included regional products.[32] Moreover, some questions about general dietary habits were included in the questionnaire and were used to adjust the responses to the FFQ following the methodology described in Calvert *et al.*[33] The inclusion of prevalent cases might be another cause for concern since patients who survived might have a very different etiology than those who died soon after diagnosis. In addition, diet can be influenced by many external factors and patients who survive longer might have substantially modified their diet. However, results of the sensitivity analysis suggested that the use of prevalent cases might not have introduced selective survival bias or reverse causation. We may have been limited by the small sample size and lack of statistical power to detect significant associations when evaluating certain subgroups. Finally, although we adjusted for a range of potential confounders, residual confounding factors cannot be totally ruled out.

The strengths of the study include the substantial sample size of CLL cases, with specific information on clinical presentation. We were able to collect detailed information on demographics and disease stage, and statistically adjust for a number of potential confounding factors. This allowed the evaluation of potential interactions of diet with numerous covariates and the exploration of the associations by stage. Finally, the multi-centric nature of the study, including both rural and urban areas, provided a wide geographic variability of dietary intake data.

In conclusion, in this Spanish population-based case-control study, greater adherence to a Western dietary pattern was associated with CLL. These novel results suggest that a proportion of CLL cases could be prevented by modifying dietary patterns. Further research, especially with a prospective design, is warranted to confirm these findings.

Acknowledgments

The authors would like to thank all the subjects who participated in the study and all CLL MCC-Spain collaborators (the list can be found the Online Supplementary Appendix, List S1).

Funding

Predoctoral contract to MS (CIBERESP), Spanish Ministry of Economy and Competitiveness Juan de la Cierva de Incorporación grant IJCI-2014-20900. Spanish Ministry of Economy and Competitiveness - Carlos III Institute of Health cofunded by FEDER funds/European Regional Develpment Fund (ERDF) - a way to build Europe [(grants PI17/01280, PI11/01810, PI14/01219, PI11/02213, PI09/1662, PI15/00966, RCESP C03/09, RTICESP C03/10, RTIC RD06/0020/0095, RD12/0036/0056, Rio Hortega CM13/00232, SV-09-CLINIC-1 and Centro de Investigación Biomédica en Red: Epidemiología y Salud Pública (CIBERESP))] and the Agència de Gestió d'Ajuts Universitaris i de Recerca AGAUR (2017SGR1085, 2014SGR756). The ICGC CLL-Genome Project was funded by Spanish Ministerio de Economía y Competitividad (MINECO) through the Instituto de Salud Carlos III (ISCIII), PMP15/00007 and Centro de Investigación Biomédica en Red: Oncología (CIBERONC). ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya.

References

1. Sant M, Allemani C, Tereanu C, et al. Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. Blood. 2010;116(19):3724-3734.
2. Slager SL, Benavente Y, Blair A, et al. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma: The InterLymph Non-Hodgkin Lymphoma Subtypes Project. JNCI Monogr. 2014; 2014(48):41-51.
3. WCRF. World Cancer Research Fund. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. 2nd ed. Washington, USA: American Institute for Cancer Research, 2007.
4. Rohrmann S, Linseisen J, Jakobsen MU, et al. Consumption of meat and dairy and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2011;128(3):623-634.
5. Daniel CR, Sinha R, Park Y, et al. Meat Intake Is Not Associated with Risk of Non-Hodgkin Lymphoma in a Large Prospective Cohort of U.S. Men and Women. J Nutr. 2012;142(6):1074-1080.
6. Tsai H-T, Cross AJ, Graubard BI, Oken M, Schatzkin A, Caporaso NE. Dietary factors and risk of chronic lymphocytic leukemia and small lymphocytic lymphoma: a pooled analysis of two prospective studies. Cancer Epidemiol Biomarkers Prev. 2010;19(10):2680-2684.
7. McCullough ML, Teras LR, Shah R, Diver WR, Gaudet MM, Gapstur SM. Artificially and sugar-sweetened carbonated beverage consumption is not associated with risk of lymphoid neoplasms in older men and women. J Nutr. 2014;144(12):2041-2049.
8. Ross JA, Kasum CM, Davies SM, Jacobs DR, Folsom AR, Potter JD. Diet and risk of leukemia in the Iowa Women's Health Study. Cancer Epidemiol Biomarkers Prev. 2002;11(8):777-781.
9. Erber E, Maskarinec G, Lim U, Kolonel LN. Dietary vitamin D and risk of non-Hodgkin lymphoma: the multiethnic cohort. Br J Nutr. 2010;103(4):581-584.
10. Chang ET, Canchola AJ, Clarke CA, et al. Dietary phytocompounds and risk of lymphoid malignancies in the California Teachers Study cohort. Cancer Causes Control. 2011;22(2):237-249.
11. Bertrand KA, Giovannucci E, Rosner BA, Zhang SM, Laden F, Birmann BM. Dietary fat intake and risk of non-Hodgkin lymphoma in 2 large prospective cohorts. Am J Clin Nutr. 2017;106(2):650-656.
12. Kabat GC, Kim MY, Wactawski-Wende J, Shikany JM, Vitolins MZ, Rohan TE. Intake of antioxidant nutrients and risk of non-Hodgkin's Lymphoma in the Women's Health Initiative. Nutr Cancer. 2012; 64(2):245-254.
13. Charbonneau B, O'Connor HM, Wang AH, et al. Trans fatty acid intake is associated with increased risk and n3 fatty acid intake with reduced risk of non-hodgkin lymphoma. J Nutr. 2013;143(5):672-681.
14. Chang ET, Smedby KE, Zhang SM, et al. Dietary factors and risk of non-hodgkin lymphoma in men and women. Cancer Epidemiol Biomarkers Prev. 2005; 14(2):512-520.
15. Chang ET, Balter KM, Torrang A, et al. Nutrient intake and risk of non-Hodgkin's

lymphoma. Am J Epidemiol. 2006; 164(12): 1222-1232.
16. Purdue MP, Bassani DG, Klar NS, Sloan M, Kreiger N. Dietary factors and risk of non-Hodgkin lymphoma by histologic subtype: a case-control analysis. Cancer Epidemiol Biomarkers Prev. 2004;13(10):1665-1676.
17. Mikhak B, Bracci PM, Gong Z. Intake of vitamins D and A and calcium and risk of non-Hodgkin lymphoma: San Francisco Bay Area population-based case-control study. Nutr Cancer. 2012;64(5):674-684.
18. Fritschi L, Ambrosini GL, Kliewer E V, Johnson KC. Dietary fish intake and risk of leukaemia, multiple myeloma, and non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev. 2004;13(4):532-537.
19. Holtan SG, O'Connor HM, Fredericksen ZS, et al. Food-frequency questionnaire-based estimates of total antioxidant capacity and risk of non-Hodgkin lymphoma. Int J cancer. 2012;131(5):1158-1168.
20. Koutros S, Zhang Y, Zhu Y, et al. Nutrients contributing to one-carbon metabolism and risk of non-Hodgkin lymphoma subtypes. Am J Epidemiol. 2008;167(3):287-294.
21. Luczynska A, Kaaks R, Rohrmann S, et al. Plasma 25-hydroxyvitamin D concentration and lymphoma risk: results of the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr. 2013;98(3):827-838.
22. Chiu BC-H, Kwon S, Evens AM, Surawicz T, Smith SM, Weisenburger DD. Dietary intake of fruit and vegetables and risk of non-Hodgkin lymphoma. Cancer Causes Control. 2011;22(8):1183-1195.
23. Casabonne D, Gracia E, Espinosa A, et al. Fruit and vegetable intake and vitamin C transporter gene (SLC23A2) polymorphisms in chronic lymphocytic leukaemia. Eur J Nutr. 2017;56(3):1123-1133.
24. Aschebrook-Kilfoy B, Ollberding NJ, Kolar C, et al. Meat intake and risk of non-Hodgkin lymphoma. Cancer Causes Control. 2012;23(10):1681-1692.
25. Campagna M, Cocco P, Zucca M, et al. Risk of lymphoma subtypes and dietary habits in a Mediterranean area. Cancer Epidemiol. 2015;39(6):1093-1098.
26. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol. 2002;13(1):3-9.
27. Erber E, Maskarinec G, Gill JK, Park S-Y, Kolonel LN. Dietary patterns and the risk of non-Hodgkin lymphoma: the Multiethnic Cohort. Leuk Lymphoma. 2009; 50(8):1269-1275.
28. Ollberding NJ, Aschebrook-Kilfoy B, Caces DBD, Smith SM, Weisenburger DD, Chiu BC-H. Dietary patterns and the risk of non-Hodgkin lymphoma. Public Health Nutr. 2014;17(7):1531-1537.
29. Castano-Vinyals G, Aragones N, Perez-Gomez B, et al. Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. Gac Sanit. 2015;29(4):308-315.
30. Castello A, Pollan M, Buijsse B, et al. Spanish Mediterranean diet and other dietary patterns and breast cancer risk: case-control EpiGEICAM study. Br J Cancer. 2014;111(7):1454-1462.
31. Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood. 2008; 111(12):5446-5456.
32. Garcia-Closas R, Garcia-Closas M, Kogevinas M, et al. Food, nutrient and heterocyclic amine intake and the risk of bladder cancer. Eur J Cancer. 2007; 43(11):1731-1740.
33. Calvert C, Cade J, Barrett JH, Woodhouse A. Using cross-check questions to address the problem of mis-reporting of specific food groups on Food Frequency Questionnaires. UKWCS Steering Group. United Kingdom Women's Cohort Study Steering Group. Eur J Clin Nutr. 1997;51(10):708-712.
34. Tablas de composición de alimentos del CESNID [Internet]. [cited 2017 Nov 9]. Available from: http://www.sennutricion.org/es/2013/05/13/tablas-de-composicin-de-alimentos-del-cesnid
35. Castello A, Buijsse B, Martin M, et al. Evaluating the Applicability of Data-Driven Dietary Patterns to Independent Samples with a Focus on Measurement Tools for Pattern Similarity. J Acad Nutr Diet. 2016; 116(12):1914-1924.e6.
36. Chen GC, Lv DB, Pang Z, Liu QF. Fruits and vegetables consumption and risk of non-Hodgkin's lymphoma: a meta-analysis of observational studies. Int J Cancer. 2013; 133(1):190-200.
37. Caini S, Masala G, Gnagnarella P, et al. Food of animal origin and risk of non-Hodgkin lymphoma and multiple myeloma: A review of the literature and meta-analysis. Crit Rev Oncol Hematol. 2016; 100:16-24.
38. Anderson JR, Armitage JO, Weisenburger DD. Epidemiology of the non-Hodgkin's lymphomas: distributions of the major subtypes differ by geographic locations. Non-Hodgkin's Lymphoma Classification Project. Ann Oncol. 1998;9(7):717-720.
39. Wu S-J, Huang S-Y, Lin C-T, Lin Y-J, Chang C-J, Tien H-F. The incidence of chronic lymphocytic leukemia in Taiwan, 1986-2005: a distinct increasing trend with birth-cohort effect. Blood. 2010; 116(22):4430-4435.
40. Clarke CA, Glaser SL, Gomez SL, et al. Lymphoid malignancies in U.S. Asians: incidence rate differences by birthplace and acculturation. Cancer Epidemiol Biomarkers Prev. 2011;20(6):1064-1077.
41. Paradis AM, Godin G, Perusse L, Vohl MC. Associations between dietary patterns and obesity phenotypes. Int J Obes (Lond). 2009;33(12):1419-1426.
42. Benavente Y, Casabonne D, Costas L, et al. Established and suggested exposures on CLL/SLL etiology: Results from the CLL-MCC-Spain study. Cancer Epidemiol. 2018;52:106–111.
43. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med. 2009;1(6):6ra14.
44. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157(1):121-141.
45. Barbaresko J, Koch M, Schulze MB, Nothlings U. Dietary pattern analysis and biomarkers of low-grade inflammation: a systematic literature review. Nutr Rev. 2013;71(8):511–527.
46. Thorburn AN, Macia L, Mackay CR. Diet, metabolites, and "western-lifestyle" inflammatory diseases. Immunity. 2014 Jun;40(6):833–842.
47. Caligaris-Cappio F. Inflammation, the microenvironment and chronic lymphocytic leukemia. Haematologica. 2011; 96(3):353–355.
48. Castello A, Fernandez de Larrea N, Martin V, et al. High adherence to the Western, Prudent, and Mediterranean dietary patterns and risk of gastric adenocarcinoma: MCC-Spain study. Gastric Cancer. 2018;21(3):372-382.
49. Castello A, Boldo E, Perez-Gomez B, et al. Adherence to the Western, Prudent and Mediterranean dietary patterns and breast cancer risk: MCC-Spain study. Maturitas 2017;103:8-15.
50. Castelló A, Amiano P, Fernández de Larrea N, al. Low adherence to the Western and high adherence to the Mediterranean dietary patterns could prevent colorectal cancer. Eur J Nutr. 2018 Mar 26. [Epub ahead of print]

Supplementary material

Table S1. Composition of food groups on the food frequency questionnaire of the MCC-Spain study and component loading for each pattern identified in the EpiGEICAM study [30].

Food group	Food	L_W[1]	L_P[1]	L_M[1]
High-fat dairy	Whole-fat milk, condensed milk, whole-fat yogurt, semi-cured, cured, or creamy cheese, blue cheese, custard, milk shake, ice-cream, double cream.	0.60	−0.11	0.20
Low-fat dairy	Semi-skimmed and skimmed milk, soy milk, skimmed yogurt, curd, cottage or fresh white cheese.	−0.49	0.60	−0.01
Eggs	Eggs.	0.19	0.08	0.16
White meat	Chicken, rabbit and duck.	0.08	0.17	0.18
Red meat	Pork, beef, lamb, liver (beef, pork or chicken), entrails, hamburgers (pork or beef) and meatballs (pork or beef).	0.27	0.09	0.22
Processed meat	Sausages, serrano ham and other cold meat, bacon, pâté, foie-gras.	0.36	0.10	0.26
White fish	Fresh or frozen white fish (hake, sea bass, sea bream), ½·salted fish and ½·smoked fish.	0.01	0.24	0.34
Oily fish	Fresh or frozen blue fish (tuna, swordfish, sardines, anchovies, salmon), canned fish, ½·salted fish and ½·smoked fish.	0.05	0.24	0.44
Seafood/shellfish	Clams, mussels, oysters, squid, cuttlefish, octopus, prawn, crab, shrimp and similar products.	0.17	0.27	0.35
Leafy vegetables	Spinach, chard, lettuce and other leafy vegetables.	−0.11	0.34	0.40
Fruiting vegetables	Tomato, eggplant, zucchini, cucumber, pepper, artichoke and avocado.	0.00	0.36	0.45
Root vegetables	Carrot, pumpkin and radish.	0.05	0.35	0.44
Other vegetables	Cooked cabbage, cauliflower or broccoli, onion, green beans, asparagus, mushrooms, corn, garlic, gazpacho, vegetable soup and other vegetables.	−0.04	0.40	0.42
Legumes	Peas, lentils, chickpeas, beans and broad beans.	0.21	0.15	0.34
Potatoes	Roasted or boiled potatoes and sweet potatoes.	0.17	0.25	0.40
Fruits	Orange, grapefruit, mandarin, banana, apple, pear, grapes, kiwi, strawberries, cherries, peach, figs, melon or watermelon, prunes, mango and papaya and other fresh or dried fruits.	−0.07	0.31	0.31
Nuts	Almonds, peanuts, pine nuts, hazelnut	0.18	0.22	0.29

Refined grains	White-flour bread, rice, pasta	0.37	0.15	0.23
Whole grains	Whole-grain bread and breakfast cereals	−0.43	0.47	−0.06
Olives and vegetable oil	Olives, added olive oil to salads, bread and dishes, other vegetable oils (sunflower, corn, and soybean).	0.12	0.19	0.34
Other edible fats	Margarine, butter and lard.	0.22	0.02	0.11
Sweets	Chocolate and other sweets, cocoa powder, plain cookies, chocolate cookies, pastries (croissant, donut, cake, pie or similar)	0.35	0.18	0.05
Sugary	Jam, honey, sugar and fruit in sugar syrup.	0.24	0.05	0.00
Juices	Tomato juice, freshly squeezed orange juice, juice (other than freshly squeezed)	0.25	0.67	−0.39
Caloric drinks	Sugar-sweetened soft drinks and nut milk.	0.74	0.21	−0.25
Convenience food and sauces	Croquette, fish sticks, dumplings, kebab, fried potatoes, crisps, pizza, instant soup, mayonnaise, tomato sauce, hot sauce, ketchup and other sauces.	0.47	0.12	0.24

[1]Component loadings for the W: Western; P: Prudent; M: Mediterranean dietary patterns.

Table S2: Association between adherence to dietary patterns and risk of chronic lymphocytic leukemia by time from diagnosis to interview, in the MCC-Spain study.

		<1 year from diagnosis		≥ 1 year from diagnosis		
	Controls N(%) (n=1.543)	**Cases N(%) (n=98)**	**OR[1] (95% CI)**	**Cases N(%) (n=271)**	**OR[1] (95% CI)**	**p-het[2]**
Western						
Q1	402 (25)	23 (23)	1	54 (20)	1	
Q2	401 (25)	20 (20)	0.89 (0.48;1.66)	70 (26)	1.30 (0.88;1.93)	
Q3	401 (25)	18 (18)	0.78 (0.41;1.49)	70 (26)	1.36 (0.92;2.02)	
Q4	401 (25)	37 (38)	1.69 (0.96;3.00)	77 (28)	1.56 (1.05;2.31)	
p-trend			0.073		0.03	
1-SD incr.			1.16 (0.93;1.45)		1.18 (1.03;1.36)	0.90
Prudent						
Q1	402 (25)	26 (27)	1	54 (20)	1	
Q2	401 (25)	26 (27)	0.90 (0.51;1.60)	66 (24)	1.13 (0.76;1.67)	
Q3	401 (25)	20 (20)	0.67 (0.37;1.24)	69 (25)	1.20 (0.81;1.77)	
Q4	401 (25)	26 (27)	0.84 (0.47;1.49)	82 (30)	1.42 (0.96;2.08)	
p-trend			0.402		0.07	
1-SD incr.			0.94 (0.76;1.16)		1.14 (0.99;1.32)	0.11
Med						
Q1	402 (25)	23 (23)	1	59 (22)	1	
Q2	401 (25)	24 (24)	0.92 (0.51;1.67)	64 (24)	0.99 (0.67;1.46)	
Q3	401 (25)	32 (33)	1.13 (0.65;1.99)	73 (27)	1.10 (0.75;1.61)	
Q4	401 (25)	19 (19)	0.65 (0.35;1.22)	75 (28)	1.13 (0.77;1.65)	
p-trend			0.313		0.43	
1-SD incr.			0.94 (0.76;1.15)		1.05 (0.92;1.21)	0.33

OR, odds ratio; 95% CI, 95% confidence interval; Q, quartile; SD, standard deviation; Med, Mediterranean; incr, increase

[1]Multinomial-logistic regression models adjusted for age, sex, level of education, energy intake (kcal/day), body mass index (kg/m^2) with province of residence as a random effect.

[2]P-value for the heterogeneity of effects

In bold: P-trend< 0.05

Table S3: Association between adherence to dietary patterns and risk of chronic lymphocytic leukemia according to treatment prior to interview, in the MCC-Spain study.

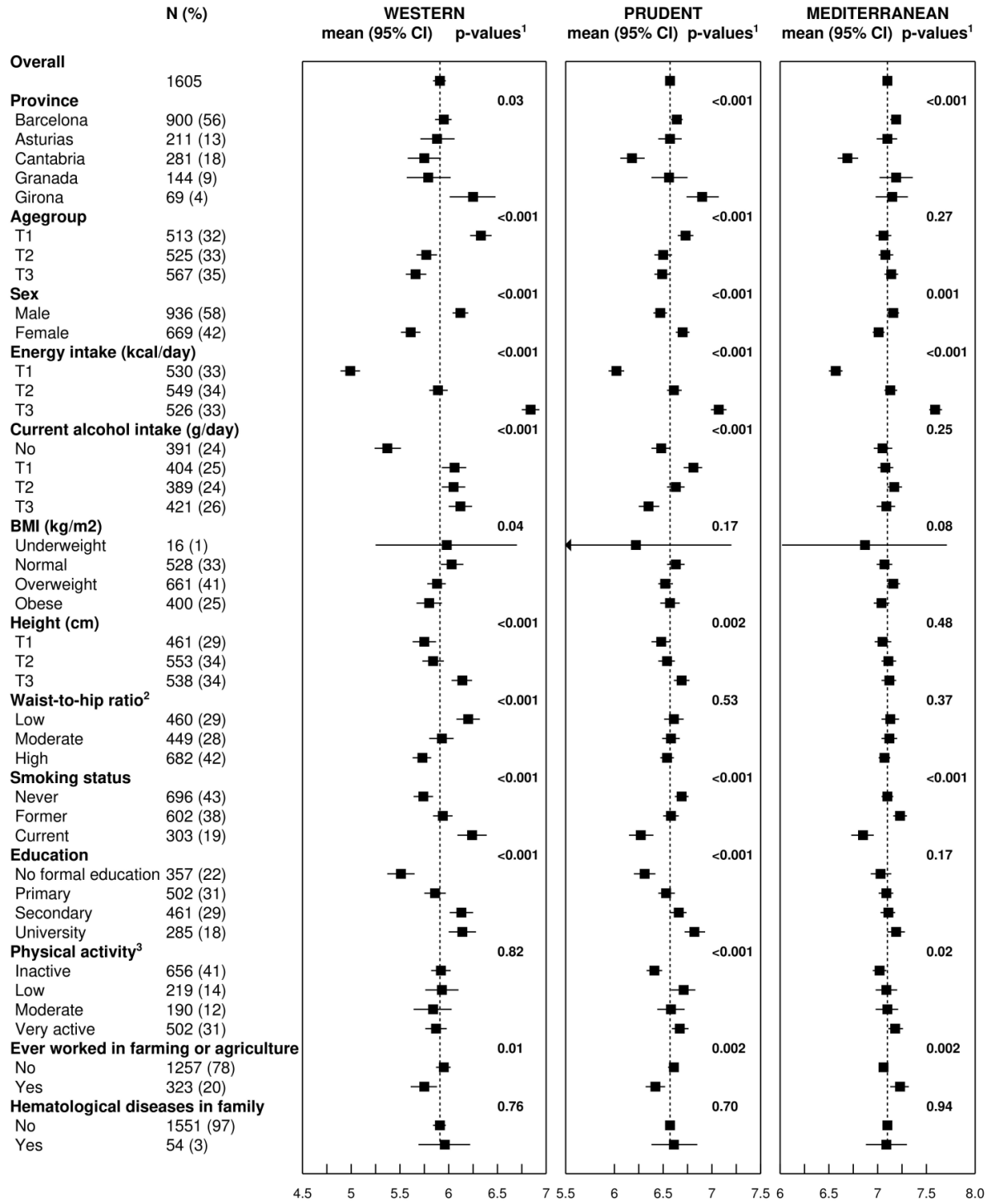
		Cases not treated before the interview		Cases treated before the interview		
	Controls N(%) (n=1,605)	**Cases N(%) (n=288)**	**OR[1] (95% CI)**	**Cases N(%) (n=79)**	**OR[1] (95% CI)**	**p-het[2]**
Western						
Q1	402 (25)	66 (23)	1	11 (14)	1	
Q2	401 (25)	71 (25)	1.11 (0.76;1.62)	18 (23)	1.54 (0.71;3.36)	
Q3	401 (25)	66 (23)	1.07 (0.72;1.60)	22 (28)	1.85 (0.85;4.03)	
Q4	401 (25)	85 (30)	1.47 (0.96;2.26)	28 (35)	**2.35 (1.04;5.31)**	
p-trend			0.11		**0.04**	
1-SD incr.			1.13 (0.96;1.33)		1.37 (1.02;1.82)	0.25
Prudent						
Q1	402 (25)	62 (22)	1	17 (22)	1	
Q2	401 (25)	72 (25)	1.02 (0.70;1.50)	19 (24)	1.06 (0.54;2.11)	
Q3	401 (25)	75 (26)	1.07 (0.72;1.57)	14 (18)	0.75 (0.35;1.59)	
Q4	401 (25)	79 (27)	1.08 (0.71;1.63)	29 (37)	1.45 (0.72;2.92)	
p-trend			0.69		0.41	
1-SD incr.			1.02 (0.88;1.19)		1.20 (0.91;1.58)	0.32
Med						
Q1	402 (25)	64 (22)	1	18 (23)	1	
Q2	401 (25)	74 (26)	0.99 (0.68;1.44)	13 (16)	0.70 (0.34;1.46)	
Q3	401 (25)	83 (29)	1.04 (0.72;1.52)	22 (28)	1.06 (0.54;2.06)	
Q4	401 (25)	67 (23)	0.79 (0.52;1.20)	26 (33)	1.15 (0.57;2.32)	
p-trend			0.34		0.47	
1-SD incr.			0.94 (0.81;1.09)		1.10 (0.83;1.45)	0.33

OR, odds ratio; 95% CI, 95% confidence interval; Q, quartile; SD, standard deviation; Med, Mediterranean, incr, increase

[1]Multinomial-logistic regression models adjusted for age, sex, level of education, energy intake (kcal/day), body mass index (kg/m^2) with province of residence as a random effect.

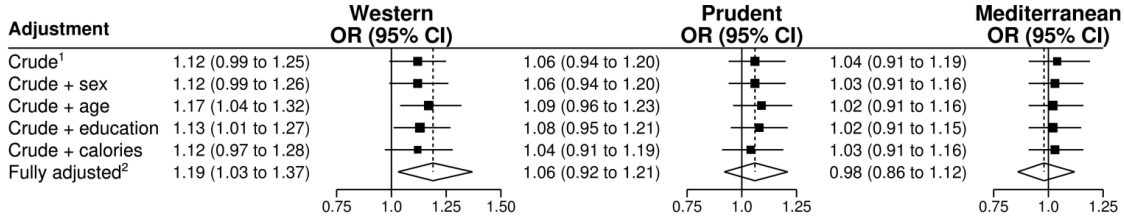
[2]P-value for the heterogeneity of effects

In bold: P-trend< 0.05



T, tertile; BMI, body mass index. Numbers do not always add up due to missing data [1]P value for heterogeneity [2]Waist to hip ratio [3]Waist to hip ratio risk categories according to WHO criteria. [4]Physical activity, in the last 10 years, measured in METs/week: inactive (0), low (0.1-8), moderate (8-15.9) and very active (≥16).

Figure S1. Means and 95% confidence intervals (CI) of levels of adherence to the dietary patterns according to characteristics of controls of the MCC-Spain study.



OR, odds ratio; CI, confidence interval.[1]Crude model includes only province as random effects. [2]Model adjusted for age, sex, education, energy intake with province of residence as random effects.

Figure S2. Odds ratios of chronic lymphocytic leukemia for a 1-SD increase in dietary patterns adjusted for one factor at a time covariate of the final model, in the MCC-Spain study.

LIST S1 – list of collaborators

MCC-SPAIN STUDY

Institut Català d'Oncologia (ICO)-L'Hospitalet de Llobregat
Víctor Moreno, Paloma Quesada, Yasmin Sabater, Marleny Vergara, Ainara, Exposito, Teresa Alonso, Isabel Padrol, Joellen Klaustermeier, Yolanda Florencia, Vanesa Camon, Eva Domingo-Domenech, Dolores Dot, Anna Esteban, Elisabeth Guinó, Santi Mercadal, Ana Oliveira, Josep Sarrà.

ISGlobal- Centre for Research in Environmental Epidemiolog (CREAL), Barcelona, Spain
Mireia García, Cecília Persavento, Esther Gracia-Lavedan

Hospital Clínic, Barcelona, Spain
Cristina Capdevila Lozar, Ainara Expósito, Silvia Martin Román, Amparo Muñoz, Yolanda Torralba

Hospital del Mar, Barcelona, Spain
Eugènia Abella, Estela Carrasco, Judith Cirac, Francesc Garcia, Antonio Salar

Institut de Prestacions d'Assistència Mèdica al Personal Municipal (PAMEM) & Centros de Atención Primaria (CAP), Barcelona, Spain
Jesús Almeda, Marifé Alvarez Rodriguez, Alex Bassa Massanas, Albert Boada Valmaseda, Enric Duran, Olga Gonzalez Ferrer, Clara Izard, Manoli Liceran, Carmen López, Josep Manuel Benítez, Dolors Petitbó, Angelina Potrony, Sònia Sarret, Laura Sebastián, Josep M. Vilaseca

Instituto Universitario de Oncología, Universidad de Oviedo, Asturias, Spain
Cristina Arias, Guillermo Fernández-Tardón, Ana Fernández Somoano, Miguel García-Villarino, Carlos López-Otín,

Servicio de Salud del Principado de Asturias, SESPA, Asturias, Spain
Enrique Colado, Begoña Martínez-Argüelles, Manuel Rivas del Fresno, Marta María Rodríguez-Suárez.

Girona group
Unitat d'Epidemiologia i Registre de Cancer de Girona: Loreto Vilardell, Montse Puig-Vives, Gemma Osca-Gelis, M Carme Carmona-Garcia, Rocio Rodriguez Romanos, Carlota Torner Galindo, Patricia Martí Bargalló, Esther Rodriguez Sanchez, Marta Solans Margalef, Aina Roca Barceló, Raquel Comas Navarro, Marc Saez
Hospital Universitari de Girona Dr. Josep Trueta: Josep Maria Roncero, David Gallardo, Rosa Coll, Ignacio Blanco.
CAP de Santa Clara: Conxa Bou
CAP de Angles: Gabriel Coll de Tuero, Alba Coll Negre

Granada group
University of Granada: Olmedo-Requena R, Olvera Porcel MC, Salcedo-Bellido I, Barrios-Rodríguez R, Jiménez-Moleón JJ;
Virgen de las Nieves,Universitary Hospital,Unit of Hematology: García-Martín P, Jurado-Chacón M.

Cantabria group
Inés Gómez-Acebo, Pilar González Echezarreta, Maria del Mar González Martínez, Luis Mariano López López, Almudena de la Pedraja Pavón, Paula Picón Sedano, Trinidad Dierssen-Sotos

Gipuzkoa-MCC group & Institutions
Pilar Amiano, Madalen Oribe, Mikel Azpiri, Ana Jimenez, Carmen Urtiaga, M. Jesús Mitxelena, Concepcion Mugarza, Irune Ruiz
Public Health Division of Gipuzkoa, Hospital Universitario Nª Sra de Aranzazu, Basque Health Department; Onkologikoa, Instituto Oncologico de Gipuzkoa

6. DISCUSSION

The research undertaken in this thesis aimed to investigate the role of several dietary patterns on risk of lymphoid neoplasms. The following sections provide a global discussion of the main results, along with their potential limitations, strengths, and future research directions.

6.1 Contribution to current knowledge

6.1.1 Comparison with previous studies on lymphoid neoplasms

To the best of our knowledge, only one prospective[91] and five case-control[41,92–95] studies have been published on this topic (**Table 5**). Overall, previous evidence mainly arises from studies with a retrospective design and generally unpowered to find subtype-specific associations. In this thesis we provided, for the first time, prospective data on *a priori* dietary patterns and lymphoma using a large study that allowed subtype analyses. In addition, results from the MCC-Spain further complement prior data on *a posteriori* patterns using the largest series of CLL/SLL cases.

Our work on the Mediterranean diet (**Paper I**) is the first prospective study to examine the association between *a priori* Mediterranean score and lymphoma risk. Consequently, results can only be tentatively compared to an Italian case-control study[41], which evaluated the association between the rMED score and lymphoma risk. The authors reported null results for overall lymphoma and an inverse trend for DLBCL, but their findings require cautious interpretation due to several important limitations: its retrospective nature, the use of modified rMED score (which did not include essential components such as olive oil, white meat or dairy products), and its limited power to detect overall or subtype-specific associations (only 322 lymphoma– and 6 DLCBL – cases were included in the analysis). Other studies have empirically derived dietary patterns presenting key features of the Mediterranean diet, and assessed its impact on lymphoma etiology. The only prior prospective study on this topic[91] reported an inverse association between a

pattern rich in vegetables and fruits and NHL (among Caucasian women), but not for other specific NHL subtypes, while case-control studies reported null results for HL[92] and NHL or its subtypes[93]. In line with this data, our results from the MCC-Spain study (**Paper III**) neither support an association between a Mediterranean-like or Prudent patterns and CLL/SLL. However, given the heterogeneity in types of foods eaten within these healthy-like patterns, the range and absolute amounts of food intakes and cut-offs used to define adherence, direct comparison of results should be made with caution. Overall, our findings suggest that adherence to a Mediterranean diet is inversely modestly associated with overall lymphoma, but not with specific lymphoma subtypes. The effect may be not strong enough to be consistently detected across studies, given the heterogeneity of definitions of dietary patterns, and the potential heterogeneity in etiology of lymphoma subtypes.

Lack of indexes to model unhealthy dietary patterns, such as a Western-like diet, has prompted that all evidence available on this topic arise from diverse empirically derived patterns. Prospective data from the US Multhiecnic Cohort[91], revealed an association between a 'Fat and Meat' pattern and FL, but not overall NHL. However, the pattern did not include sweets and deserts, sweetened beverages, or convenience foods, which may be important contributing factors influencing lymphomagenesis. By contrast, the two case-controls studies published on this topic did reveal positive associations with NHL[93] and HL[92]. Concretely, Ollberding *et al.*[93] pointed out that a high adherence to a 'Meat, Fat and Sweets' dietary pattern, characterized by a high intake of French fries, red meat, processed meat, pizza, salty snacks, sweets and desserts, was associated with an increased risk of overall NHL in a Nebraska case-control study. This association was maintained when stratifying according to lymphoma subtypes (FL, DLBCL and marginal zone lymphoma), but not for CLL/SLL, T-cell or other miscellaneous B-cell lymphoma – probably due to small sample size (e.g. CLL/SLL n= 25). Similarly, Epstein *et al.*[92] reported a strong association between a diet rich in sweets and desserts and overall HL, and between a dietary pattern rich in red and processed meat and HL in cases aged ≥50 years. Complementing this findings, we provided, for the first time, evidence of a strong association between a Western-like pattern

and CLL/SLL (**Paper III**). Again, not only differences in the study design and setting, but also in food groups loaded in these data-driven analyses, should be carefully considered when comparing results. Altogether, these findings support cumulative evidence pointing to a link between a diet rich in fats, meat (namely red and processed meat), sweets, and convenience foods and lymphoid neoplasms risk.

We hypothesize that the link between a Western-like diet and lymphoid neoplasms could partly explain higher rates of lymphoma in Western countries[8] and the dramatic increase of NHL incidence 1970's until the 20th century[11]. Changing lymphoid incidence patterns with immigration and acculturation[96,97,113] further support this hypothesis. For example, a dramatic increase in CLL incidence rates in Taiwan has been associated with a strong birth-cohort effect, that corresponded to the westernization of lifestyle since 1960[113]. Similarly, higher CLL incidence rates have been reported among US-born Asians compared to foreign-born Asians[97]. Additional studies on *a priori* dietary patterns clearly defining a Western diet, or the analysis of emerging scores quantifying ultraprocessed food consumption[114] could help to clarify this association.

Indexes on the inflammatory potential of diet offer an additional perspective on the effect of overall diet on lymphoid neoplasms etiology. We provided the first prospective data assessing the impact of a pro-inflammatory diet and risk of lymphoma and its subtypes (**Paper II**). Our findings are in line with those reported in a multicenter case-control Italian study, which found an association between the DII and overall NHL (stronger among men) and DLBCL[94], and null results for HL[95]. However, although the DII and ISD have been shown to highly correlate in the EPIC population (Pearson's correlation coefficient: 0.91; p-value<0.001)[88], data from both studies cannot be directly compared with ours, since they are based upon different indexes and study designs. In addition, the Italian case-control study lacked information on potential confounders (e.g. BMI or physical activity) as well as on NHL entities other than DLBCL or FL. So far, the body of evidence suggests that adherence pro-inflammatory diet is associated with NHL – particularly B-cell lymphomas. Given that a Western dietary pattern has been related to higher levels

of inflammatory biomarkers[115–117], these findings further sustain the potential role of a Western diet in lymphomagenesis mediated through inflammatory processes.

Interestingly, we observed strong associations between both the arMED and ISD scores and HL albeit statistically non-significant (**Papers I and II**). Despite the large number of enrolled subjects at baseline and long follow-up, the number of observed incident HL was low (n=135). Therefore, the study might not have sufficient power to detect significant associations within this subgroup. Only one previous (case-control) study has studied dietary patterns in HL[92]. With a larger set of HL cases (n=435 classical HL), that allowed exploratory analyses by age group (<50 or ≥50 years), histological subtype (nodular sclerosis or mixed cellularity) and EBV status (positive or negative), the authors observed positive associations between adherence to two dietary patterns ('Sweets/Desserts' and 'Western style') with apparent variation by age group and tumor EBV status. Together, these and our results suggest that HL might be a lymphoma prone to be influenced by dietary patterns. Thus, further prospective studies with detailed dietary data and a larger set of HL cases, probably only feasible through pooling data from consortium studies, are warranted.

6.1.2 Comparison with previous studies on solid neoplasms

The dietary patterns examined in the current thesis have been previously studied in solid neoplasms across the EPIC and MCC-Spain studies (**Table 10**)**.** The strength of the association between a Western diet and CLL/SLL was similar to that observed for breast[118], colorectal[119] and gastric[120] cancers in the MCC-Spain study. Unlike CLL/SLL, those sites, as well as aggressive prostate cancer[121], have also shown noteworthy inverse associations with the Mediterranean-like dietary pattern. EPIC investigators have also reported inverse associations between a Mediterranean diet and breast[73], colorectal[122], and gastric[106] cancers (but not with bladder[123] and pancreatic[124] sites), which seem to be stronger than results for lymphoma. In the same vein, the link between the ISD and lymphoma was more modest than that found for gastric[88] cancer in the same study. Overall, the effect of

dietary patterns in cancer risk seems to be stronger for solid neoplasms (at least in breast, colorectal, and gastric cancers) than for lymphoid neoplasms. And within lymphoma, detrimental dietary patterns tend to be more consistently associated with lymphoma etiology rather than healthy-like ones.

Table 10: Association between dietary patterns and different cancer sites within the same study populations.

MCC-Spain study (*a posteriori* patterns)	**Western** OR_{Q4vsQ4} (95% CI)	**Prudent** OR_{Q4vsQ4} (95% CI)	**Mediterranean** OR_{Q4vsQ4} (95% CI)
CLL/SLL (**Paper III**)	1.63 (1.11; 2.39)	1.16 (0.80; 1.67)	0.89 (0.61; 1.29)
Breast cancer[118]	1.53 (1.15; 2.02)	0.92 (0.70; 1.20)	0.90 (0.69; 1.17)
Premenopausal	1.68 (1.02; 2.79)	1.00 (0.65; 1.53)	1.39 (0.92; 2.11)
Postmenopausal	1.48(1.07; 2.05)	0.89 (0.65; 1.21)	0.72 (0.53; 0.98)
Colorectal cancer[119]	1.50 (1.20; 1.87)	0.94 (0.76; 1.15)	0.65 (0.53; 0.80)
Proximal colon	1.19 (0.85; 1.66)	0.92 (0.67; 1.28)	0.70 (0.51; 0.97)
Distal colon	2.02 (1.44; 2.84)	1.06 (0.78; 1.44)	0.65 (0.48; 0.89)
Rectum	1.46 (1.05; 2.01)	0.83 (0.62; 1.12)	0.60 (0.45; 0.81)
Prostate cancer[121]	1.15 (0.83; 1.58)	0.94 (0.69; 1.28)	0.90 (0.66; 1.23)
Non-aggressive[a]	1.18 (0.78; 1.81)	1.29 (0.85; 1.97)	1.31 (0.86; 1.99)
Aggressive[a]	1.11 (0.75; 1.65)	0.78 (0.54; 1.14)	0.68 (0.46; 1.01)
Gastric cancer[120]	2.09 (1.31; 3.33)	1.40 (0.93; 2.11)	0.53 (0.34; 0.82)
Cardia	1.40 (0.66; 2.99)	1.58 (0.73; 3.40)	0.74 (0.30; 1.86)
Non-cardia	2.01 (1.21; 3.35)	1.32 (0.82; 2.12)	0.47 (0.29; 0.77)
Intestinal	2.74 (1.38; 5.45)	1.55 (0.83; 2.87)	0.57 (0.29; 1.12)
Diffuse	2.10 (0.86; 5.10)	1.74 (0.76; 3.99)	0.90 (0.38; 2.15)

EPIC study (*a priori* scores)	**ISD** HR 1-SD (95% CI)	**Mediterranean diet**[b] $HR_{highvslow}$ (95% CI)
Lymphoma (**Papers I & II**)	1.05 (1.00; 1.11)	0.91 (0.80; 1.03)[c]
HL	1.22 (0.94; 1.57)	0.64 (0.34; 1.19)
NHL	1.06 (1.00; 1.13)	0.94 (0.82; 1.08)
Mature T/NK-cell	0.99 (0.76; 1.29)	0.78 (0.42; 1.44)
Mature B-cell	1.07 (1.01; 1.14)	0.95 (0.82; 1.10)
Breast cancer[73]	-	0.94 (0.88; 1.00)
Premenopausal	-	0.97 (0.81; 1.15)
Postmenopausal	-	0.93 (0.87; 0.99)
Colorectal cancer[122]	-	0.89 (0.80; 0.99)
Gastric cancer[88,106]	1.25 (1.12; 1.39)	0.67 (0.47; 0.94)
Cardia	1.30 (1.06; 1.59)	0.45 (0.21; 0.91)
Non-cardia	1.07 (0.89; 1.28)	0.71 (0.44; 1.17)
Intestinal	1.18 (1.03; 1.34)	0.61 (0.34, 1.11)
Diffuse	1.33 (1.06; 1.67)	0.69 (0.39, 1.22)
Bladder cancer[123]	-	0.84 (0.69, 1.03)
Non-aggressive	-	0.78 (0.54, 1.14)
Aggressive	-	0.88 (0.61, 1.28)
Pancreatic cancer[124]	-	0.99 (0.77, 1.26)

OR, odds ratio; CI, confidence interval, HR, hazard ratio; arMED, adapted relative Mediterranean diet; ISD, inflammatory score of diet; HL, Hodgkin lymphoma, NHL, non-Hodgkin lymphoma.
Results in red (positive association) or green (inverse association) indicate p-value<0.05.
[a]Results are expressed as risk ratio
[b]Scores: arMED (for lymphoma and breast cancer), rMED (for gastric, bladder, and pancreatic cancers), and Trichopoulou score (for colorectal cancer).
[c]Inverse association when assessed as a continuous score (HR 1-unit increase = 0.98 (0.97; 1.00)

6.1.3 Biological explanation

Individual dietary factors can influence carcinogenesis by affecting fundamental cellular processes, including those that regulate the balance between cell proliferation, differentiation, and death, the expression of oncogenes and tumor-suppressor genes, cell signaling, and other factors in the cellular environment, such as hormonal factors, that further influence gene expression[23,58] (**Figure 9**). Based on global basic and epidemiological research, the WCRF/AICR has proposed several mechanisms that may be mediating the associations between several dietary components (e.g. wholegrains, vegetables, fruit, meat, fish, dairy, food processing, and alcoholic drinks) and solid neoplasms[23]. Overall, foods contain numerous bioactive components which, in turn, can influence (additively or even synergistically) gene expression at multiple points. Thus, the association between dietary patterns and lymphoid neoplasms is biologically plausible but hard to disentangle.

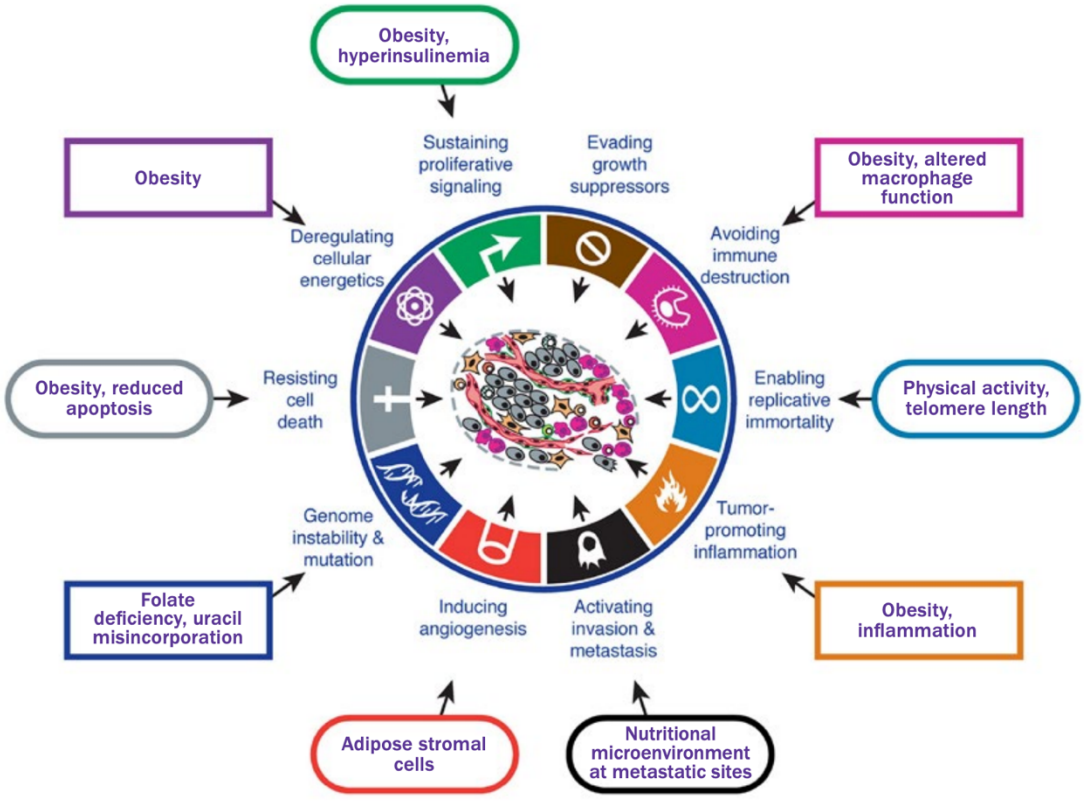


Figure 9. Nutrition, physical activity, and the hallmarks of cancer (adapted from WCRF/AICR 2018 report[23]).

More specifically for lymphoid neoplasms, it has long been suspected that lymphomagenesis may be mostly driven by perturbations of the immune system. The reason is that severe immunodeficiency (both hereditary or acquired)[125], certain autoimmune and chronic inflammatory conditions[126,127], and several infections associated with chronic immune stimulation[128], are among the best characterized and strongest known risk factors of lymphoma. For the rest – and majority – of lymphoma cases, however, the cause remains unknown, but subclinical changes in the immune system could be likewise involved. Recent studies within general population cohorts incorporating serologic measurements of cytokines, chemokines, and other immune markers support the role of subtle immunologic effects in lymphomagenesis[129–138]. Thus, in the recent years, several authors have suggested that the association between other potential risk factors, including environmental and occupational exposures, and lymphoma may be also exerted though modulation of the immune function[139]. Results from **Paper II** further support this hypothesis, suggesting that chronic inflammation induced by diet may be one of the mechanisms explaining the link between nutrition and B-cell lymphoma. Mediation analyses – specially focused on immune markers – could further elucidate the mechanisms underlying the associations found.

Similarly, an inflammation-related mechanism may also underlie the observed associations between a Western-like diet and CLL/SLL (**Paper III**), given that a diet high in fat, refined grains, red and processed meats, and sweets has been largely associated with higher levels of inflammatory markers[140] and with inflammation-related chronic diseases[141]. In addition, it is well-stablished that dietary changes, and particularly switching from a low-fat, plant polysaccharide–rich diet to a high-fat, high-sugar Western diet, can induce alterations in microbiota composition[142]. Beyond its role in the biosynthesis of key components (e.g. vitamins, essential amino acids or short chain fatty acid byproducts), several studies using germ-free mice suggest that microbiota also plays a fundamental role on the induction, training, and function of the host immune system[143]. Exposure to a Western diet may have selected for a microbiota that lack the resilience and diversity required to establish balanced immune responses, which could also contribute to explain the associations found. In addition, a wealth of evidence

further support an association between main components of a Western dietary pattern (i.e. highs amounts of free sugars, starches, fats, and processed meat) and numerous solid neoplasms[23]. Suggested mechanisms include i) pro-inflammatory effects, ii) hyperinsulinemia, insulin resistance, and enhanced bioactivity of the IGF axis, iii) carcinogenic compounds derived from cooking methods (i.e. heterocyclic amines and polycyclic aromatic hydrocarbons), or found in processed meat products (i.e. N-nitroso compounds), among others, as well as an indirect effect, by causing weight gain, overweight, and obesity, which, in turn, are well-stablished risk factors for several neoplasms. Indeed, based on these evidence, the current WCRF/AICR recommendations for cancer prevention promote limiting the consumption of fast food, red and processed meat, and sugar sweetened drinks[23]. Mechanistic evidence on these components and lymphoma is far more limited, but overall, our results suggest that a combined/synergistic/interactive effect of these components, probably mediated through multiple biologic pathways, could be contributing to the associations found.

The potential mechanisms underlying the inverse association between a Mediterranean diet and cancer risk have been largely studied[65] and would likewise be considered for lymphoma (**Paper I**). While the exact mechanism is not known, accumulating evidence point to 5 key processes[144]: i) lipid-lowering effect, ii) protection against oxidative stress and inflammation iii) modification of hormones and growth factors involved in the pathogenesis of cancer, iv) inhibition of nutrient sensing pathways by specific amino acid restriction, and v) gut microbiota-mediated production of metabolites influencing metabolic health. Overall, while many of the Mediterranean diet components provide potential physiological explanations for its beneficial effect on lymphoma risk, it is the unlikely that one individual component is entirely responsible. In line with previous studies, our results further support that the beneficial effect probably comes from a combinatory effect of many interrelated and overlapping dietary factors.

6.2 Methodological considerations: strengths and limitations

6.2.1 Study design

The strengths of the EPIC cohort (**Papers I and II**) include its prospective design, long follow-up and large sample size, which allowed us to carry out analyses by lymphoma subtypes and assess the long-term effect of several dietary exposures. On the other hand, the MCC-Spain study design (**Paper III**) was appropriate for the evaluation of an outcome with such a long induction period, providing the largest set of CLL/SLL cases in which the effect of dietary patters has been assessed. Furthermore, in both studies we were able to collect detailed information on diet, demographics and lifestyle and statistically adjust for a number of potential confounding factors. Finally, the multi-centric design of both the EPIC and MCC-Spain studies allowed the inclusion of a geographically diverse population, covering a wide range of dietary patterns and lifestyle habits through Europe and Spain, respectively. However, several limitations should be considered when interpreting the results of these studies, which are detailed in the specific papers and further addressed in the following sections.

6.2.2 Dietary data collection and measurement

Dietary data from both the MCC-Spain and EPIC studies arise from food FFQ (self-administered in the majority of centers), which assessed the frequency of intake of listed food items during the previous year[101,107]. Data derived from this instruments invariably differ from the true intake values for several reasons: subjects may find it difficult to recall and average their intakes over a long period, answers may be influenced by psychological factors (e.g. social desirability), and consumption frequencies and average proportion sizes of food groups may be imperfectly translated into specific nutrient amounts[145,146]. Thus, the degree of accuracy in measures of the exposures and other co-variables could have influenced the observed associations. However, how this measurement error influences the results depends both of its magnitude and whether the misclassification was differential or non-differential[147].

Differential (or non-random) measurement error results in errors in the exposure classification that differ within study subgroups, such as cases and non-cases. It typically occurs in case-control studies, when case subjects recall their diet with different error than control subjects, resulting in recall bias, one of the major limitations of **Paper III**. This type of error is less likely to occur in a cohort study where information has been collected from healthy subjects before the onset of the disease.

On the other hand, there is non-differential (or random) measurement error, which is uncorrelated with the disease. This type of misclassification is common in nutritional epidemiology because of the inherent difficulties of measuring dietary exposure[145]. Despite using validated country-specific questionnaires, additionally using cross-check questions on aggregated food group consumption in the MCC-Spain study, and excluding cases with implausible energy intakes later in the analyses, some degree of random measurement error is unavoidable. However, if these errors were present, they could be equal between all groups (exposed, unexposed, cases and non-cases), resulting in an attenuation of the measures of association (i.e. biasing the estimates towards a null effect)[147]. An additional approach to partially correct for non-differential dietary measurement error is to take a secondary more accurate dietary assessment, which in the case of the EPIC study was a 24-h dietary recall carried out in a sub-sample of the main cohort[148]. This reference data – taken as a gold standard – is then used to adjust for systematic over- or underestimation in dietary information obtained from the dietary questionnaires. Dietary calibration, using the linear regression calibration approach, has been shown to correct for attenuation in risk estimates as well as between-center heterogeneity within the EPIC study[149]. However, we did not repeat our analyses using calibrated dietary data in **Papers I and II,** given that it could only marginally de-attenuate our estimates which could, in turn, remain modest. Unfortunately, over- and under-estimates in the FFQ have been shown to be highly correlated with those in the 24-h dietary recall[150]. So far, biomarkers of nutrient intake represent the optimal standard for calibration of FFQs, as the errors of the two methods are uncorrelated and therefore a true reduction of this bias could be possible[146]. Biological markers are currently not available for most

nutrients, but much hope is held for their development and later link to dietary patterns[151].

In addition to measurement errors, dietary changes also pose an additional challenge to epidemiological studies. First, several individuals could potentially have symptoms that lead to dietary changes (presumably to a more healthy-like diet) before diagnosis. This might lead to erroneous diet–cancer associations (reverse causation), and could be present both in the MCC-Spain and EPIC studies. However, in the latter, this could be minimized by excluding those cases diagnosed during the first two years of follow-up (and thus, who might have pre-diagnostic symptoms when completing the FFQ) in sensitivity analyses, which did not reveal any relevant changes in our results. Finally, EPIC participants may have changed their lifestyle habits during the follow-up period. Ideally, additional dietary measurements should be made periodically throughout the follow-up to identify any important dietary changes. For our analyses, though, we have had to assume the diet measured through the FFQ was stable through the 14-year average follow-up.

6.2.3 Exposure assessment: dietary patterns' construction

There are several advantages of evaluating the effect of dietary patterns on risk of lymphoma, as opposed to the traditional approach of focusing on individual foods or nutrients (*see* **section 1.2.1**). Below, the specific strengths and limitations of the scores and patterns assessed in the current thesis are briefly described.

6.2.3.1 The arMED score

The arMED score has three advantages in relation to other a priori Mediterranean dietary scores[72]. First, it is based on energy-adjusted tertiles of intake of each component, to take into account the quantity consumed relative to an individual's energy intake (nutrient density approach)[152]. In addition, the use of sex-specific tertiles is unnecessary, given that differences in total energy intake between sexes are already considered. Secondly, the use of tertiles, in comparison to the commonly used dichotomous cut-offs, better discriminates between the variations

of intake in the study population. Thirdly, the use of olive oil instead of the monounsaturated/saturated fatty acid ratio which is a proxy for olive oil but also an indicator of meat intake (specially in non-Mediterranean countries)[153], may better capture the Mediterranean dietary pattern.

On the other hand, several limitations – generally attributable to most dietary indexes – must be considered. Constructing the arMED involves making choices regarding i) which foods or nutrients to include, ii) how components contribute to the score, iii) which cut-offs to use, and iv) how to analyze the final score (i.e. which ranges determine low, moderate, and high adherent individuals). The consequences of these choices should always be addressed, especially when they are due to arbitrary decisions[154]. More specifically, the arMED only considers whether consumption of each component is above or below a certain threshold. This one, in turn, might not be represent an optimum level of consumption, since cut-off points are not based on evidence derived from dietary guidelines but are population-dependent (tertiles). Consequently, the cut-off points very between populations and, as a result, a high adherence to a Mediterranean dietary pattern in this analysis is not directly comparable to other studies using the same scoring within different populations. A further arbitrary decision is that the index gives equal weights to its components, assuming they have equal effects on the health outcome. For instance, the meat component encompasses all types of meats, including processed meat, which has been classified by IARC as carcinogenic to humans (Group 1)[155], along with white meat, which, to the date, has not been consistently associated with cancer[156]. The same occurs between major components, such as vegetables and cereals, which might have distinct effects on health outcomes but are finally scored equally.

Finally, alcohol treatment in dietary patterns, which can have either detrimental or positive effects depending on the health outcome under study, is still a challenge[157]. The original rMED score[73] included alcohol in the scoring system: 2 points are assigned for moderate consumers (5–25 g/day for women and 10–50 g/day for men) and 0 points for those above and below the sex-specific range, owing its moderate consumption in Mediterranean populations (mainly wine with

meals) and its beneficial effects when consumed in moderation. However, while convincing evidence suggests that alcohol increases the risk of several carcinomas[23,158], accumulating evidence shows an inverse association between increasing alcohol intake and NHL, especially for DLCBL and FL. Thus, to ensure that alcohol was not driving the associations found, and in line with other publications[73], we decided not to include it in the scoring system, but adjusted and further stratified our models by alcohol intake. Further studies are needed to clarify its optimal treatment in dietary pattern analyses.

6.2.3.2 The ISD

Several strengths of the ISD, also applicable to the DII[87], must be acknowledged. First, the ISD is grounded in previous evidence from epidemiological and basic research on the effect of diet (including foods, micro and macronutrients) and inflammation. In addition, it was designed without regard to any particular health endpoint in other to reduce possible bias. Moreover, it avoids the arbitrariness resulting from using raw intakes units, by first standardizing the intake of food parameters using the EPIC population, and then calculating the centered percentiles, which also dilutes the effect of extreme values that could results in skewness of the score. Finally, and in comparison with the DII, it introduces two modifications that improve the scoring system. First, total fat is dismissed, since its inflammatory effect is likely to be represented by the weights of all separate components of fats already included in the score (i.e. saturated, mono-unsaturated, and polyunsaturated fats). Secondly, alcohol is considered to be anti-inflammatory for all levels of consumption in the DII, but this property has only been reported in literature for low/moderate consumers (less than 30-40 g/day)[159,160]. Therefore, for subjects with intake >40 g/day the weight for alcohol in the ISD was set to 0.

As far as limitations are concerned, in line with other indexes, the development of the scoring algorithm implied certain arbitrary decisions (e.g. the scoring of articles reviewed according to the study design). In addition, its standardization allows quantitatively comparisons with other EPIC publications, but at the same time, hampers comparability with studies using the DII[161]. Furthermore, the score

may be also affected by publication bias (the inclusion of significant findings is more likely than null ones), and become rapidly obsolete, since it relies on an objective knowledge base that is continuously growing. However, the scoring computation has been shown to be consistent across the last literature updates. More concretely, between the defunct version of the DII[162] and the current one[87], the authors reviewed and scored 3 additional years of peer-reviews publications (i.e. from 2007 to 2010). While the total literature size doubled during this period, the weighting of components remained consistent (i.e. nothing that was shown to be anti-inflammatory as of 2007 was found to be proinflammatory or null as of 2010, and the same for proinflammatory items)[161].

6.2.3.3 A posteriori-derived dietary patterns

Limitations of the Western, Prudent, and Mediterranean-like dietary patterns examined in this thesis (**Paper III**) are shared with other a *posteriori* dietary patterns. As briefly mentioned in **Table 3**, the PCA approach involves several arbitrary but important decisions, including the grouping of food items into food groups, the numbers of factors to extract, the method of rotation, and the naming of the selected components[56]. Regarding food groupings, for instance, the FFQ of the MCC-Spain did not specify the origin of the item "hamburger", which could be both categorized as red meat (if handmade) or processed meat (if fast food), and thus, its final categorization as "red meat", although substantiated by assumptions regarding the profile of the study population, was subjective. Overall, this potential lack of robustness hampers the reproducibility and validity of this methodology, which is one of the main limitations of this approach. However, some authors have shown that scores of adherence to dietary patterns can be calculated following the same rules over different populations, resulting in different levels of adherence while still being valid[163]. Our results, based on three dietary patterns originally extracted in the EpiGEICAM study and latter applied in the MCC-Spain study, further sustain it.

6.2.3.4 New approaches and existing gaps

The dietary pattern analysis field is rapidly evolving, with multiple scores and new methodological approaches continuously emerging. For instance, we recently proposed the application of compositional data analysis (CoDA) to extract dietary patterns, owing the compositional nature of diet. CoDA is standard family of statistical methods for analyzing the relative importance of magnitudes, which holds a great potential in the context of dietary patterns. In particular, we emphasized the use of principal balances, which overcome the limitations of purely data-driven methods and present dietary patterns as trade-offs between eating more of some foods and less of others (*see* **Annex 2**).

However, as evidenced in a workshops at the National Institutes of Health (NIH), "Extending Dietary Patterns Research Methods"[157], several gaps must be covered in order to further advance in this field. In broad terms, the panel acknowledged that dietary patterns are conceptualized and defined in many ways: "as an exposure or a behavior, by numbers or labels, as univariate or multivariate constructs, as research-driven or data-driven, and as static or dynamic". Overall, this generates too heterogeneous literature to draw firm conclusions about the effect of specific dietary patterns on health, as stated in the WCFR 2007 report[20] and in the 2015 Scientific Report of the Dietary Guidelines Advisory Committee[164]. Thus, there is a need to standardize dietary methods and scores, as well as to better model the multidimensionality (i.e. it may be not ideal to reduce all aspects of diet to a single, unidimensional diet quality metric or score) and dynamism (i.e. of dietary patterns change over time and life course) of dietary patterns.

6.2.4 Outcome assessment and statistical power

Etiological research of most lymphoma subtypes is hampered by their low incidence – in fact, most of them are considered rare cancers[165]. The study of subtype-specific associations in lymphoma thus, has traditionally been set aside for a more general approach usually focused on overall NHL due to limited sample size. Given that the associations between diet and NHL risk seem to differ by histopathologic subtype, some of the discrepancies among previous results could

be due to different proportions of these subtypes in the case populations and lack of statistical power. In this sense, one of the major strengths of this thesis are the analyses by accurately diagnosed and classified lymphoma subtypes. Concretely, the MCC-Spain study (**Paper III**), included a substantial sample size of CLL/SLL cases, diagnosed according to the International Workshop on CLL criteria, morphologically and immunologically confirmed, and additionally categorized using the Rai staging system. However, not only incident but also prevalent cases were included in the study to enlarge the sample size. This can be a cause for concern, since prevalent cases are less likely to accurately report past exposures and might have a different etiology than those who died soon after diagnosis. In addition, diet can be influenced by many external factors, and patients who survived longer might have substantially modified their diet as a result of the onset of the disease. However, we observed minimal impact on estimates by restricting the analysis to incident cases, suggesting that the use of prevalent cases might not have introduced selective survival bias or reverse causation. As far as the EPICs study is concerned (**Papers I and II**), it included incident lymphoma cases classified into the ICD-O-3, and subsequently categorized into more specific subgroups according to the InterLymph hierarchical classification[7]. However, we still may have been limited by sample size to find associations for entities, such as HL, with less than 150 cases after an average of 14 years of follow-up.

Overall, more studies with detailed classification schemes – and ideally with a prospective design – are needed to clarify the associations found. Given the very large sample sizes and long follow-up durations required to capture the development of these diseases, this might only be addressed through meta-analyses of published works and pooling of studies through consortium efforts such as InterLymph[166].

6.2.5 Dealing with confounding

Dietary patterns are closely related to other health-related behaviors[69,167–169]. In general, individuals with high quality diets are likely to follow other healthy lifestyle behaviors, while less healthful dietary patterns tend to be positively

associated with smoking, alcohol consumption, and low physical activity, among others. Thus, given that lifestyle behaviors tend to cluster together, our results could be confounded if these variables (and others, such as socioeconomic status, energy intake, age, BMI, or sex) were also associated with lymphoma risk without being on the causal pathway between exposure and outcome.

Both in the EPIC and MCC-Spain studies, there is comprehensive data on socio-demographic, anthropometric, lifestyle, and dietary factors. These allowed us to correct our models by all potential confounders available. However, residual confounding in observational studies is almost inevitable by additional confounders not taken into account or by measurement error in those assessed. While it is recognized that, under certain conditions, nondifferential measurement error in the exposure leads to a bias towards the null, the effects of measurement error in confounders are not well understood[170]. A major alternative design strategy for dealing with confounding is the randomized controlled trial: the randomization process renders groups similar with respect to both known and unknown confounders. Interventional studies provide the highest level evidence for causal relationships, although are often difficult to implement when studying chronic disease endpoints. However, studies such as PREDIMED[171] are providing a wealth of evidence supporting the role of the Mediterranean diet in preventing cardiovascular disease risk factors and, more recently, have reported a beneficial effect of a Mediterranean diet supplemented with extra-virgin oil in the primary prevention of breast cancer[172]. Similar intervention studies, however, will hardly be handled for lymphoma given its low incidence, and certainly, will never assess harmful interventions such as a 'Western-like diet'. Overall, new evidence on dietary patterns and lymphoma will probably be arising from observational nutritional studies inevitably affected by cofounding.

6.3 Public health implications and future research directions

This thesis provides the most comprehensive data on the role of dietary patterns in the etiology of lymphoid neoplasms. According to our results, the influence of these factors in lymphomagenesis seems to be moderate. Suggestive associations observed for HL warrant further evaluation of this and other low-incident subtypes in enough powered studies.

Our findings support the accruing evidence on the benefits of the Mediterranean diet on health and longevity[74]. Incorporating this knowledge to practical, regionally tailored dietary guidance and policies worldwide is crucial, owing the tendency of societies to shift toward a more westernized dietary pattern and the increasing global burden of cancer and other chronic diseases. Unfortunately, this also applies to Mediterranean countries, in which diet is shifting away from the traditional pattern and is increasingly including meat and fat of animal origin[173].

With the exception of studies examining the inflammatory potential of diet, most of the evidence on "unhealthy" diets arise from *a posteriori*-derived dietary patterns, which are difficult to compare and synthesize. It could be insightful to investigate other reproducible dietary scores, such as the glycemic index/load[174], which ranks carbohydrate foods according to their ability to raise blood glucose levels, or the NOVA classification[114], which categorizes products according to their degree of food processing. Not only would they provide additional evidence on the link of diet and lymphoma, but also clarify the potential physiological mechanism behinds its effects.

Mechanistic studies will also be needed to shed light on the underlying biology of how these dietary patterns may modulate lymphoma disease initiation (and maybe progression). Given that subclinical changes in the immune system seem to have an important role in lymphomagenesis, studies relating dietary patterns with immune biomarkers are a particularly relevant area of future research[139].

Finally, several studies have constructed combined diet and lifestyle scores, such as the Healthy lifestyle index (HLI)[175] or the WCRF/AICR score[85]. Whilst having been extensively studied for solid neoplasms, these exposures remain unexplored in lymphoma. A multidimensional lifestyle approach exploring the joint effect of diet and other lifestyle behaviors, such as level of physical activity, alcohol consumption or smoking status, would provide additional information for clarifying disease etiology.

Overall, pooled analyses through consortia will be needed to investigate more thoroughly the associations between lymphoma and dietary factors. Results of these studies should substantially advance our understanding of the link between diet and lymphoma risk, which should ultimately be translated into prevention programs aimed at reducing the public health burden of lymphoma worldwide.

7. CONCLUSIONS

Specific conclusions:

Results from the EPIC study (papers I and II):

- A greater adherence to a Mediterranean diet is modestly associated with decreased risk of overall lymphoma, but not with specific subtypes.
- A pro-inflammatory diet is modestly associated with mature B-cell neoplasms risk, but not with any specific subtype within this group.
- HL might be a lymphoma subtype prone to be influenced by dietary pattern exposition, and this association could be clarified in further prospective studies including a larger subset of HL cases.

Results from the MCC-Spain study (paper III):

- A Western-like dietary pattern is strongly associated with CLL/SLL, independently of Rai stage at diagnosis.
- A Prudent and Mediterranean-like dietary patterns are not associated with CLL/SLL.

Final comments:

- Dietary patterns are likely to play a modest role in lymphomagenesis, yet the biological mechanisms involved need to be elucidated.
- Their effect may be not strong enough to be consistently detected across studies with a different case mix, given the heterogeneity of definitions of dietary patterns and the heterogeneity in the etiology of lymphoma subtypes.
- Further studies – ideally with a prospective design and with subtype-specific data – are warranted to confirm these findings. Given the low incidence of most subtypes, pooled analyses through consortia will be needed to investigate more thoroughly the associations between lymphoma and overall diet.

REFERENCES

1. Jaffe ES, Harris NL, Stein H, Isaacson PG. Classification of lymphoid neoplasms: the microscope as a tool for disease discovery. *Blood* 2008;112:4384–99.
2. Morton LM, Turner JJ, Cerhan JR, Linet MS, Treseler PA, Clarke CA, Jack A, Cozen W, Maynadié M, Spinelli JJ, Costantini AS, Rüdiger T, et al. Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). *Blood* 2007;110:695–708.
3. Swerdlow SH, Campo E, Harris NL, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC Press; 2008. 2001. 171–175p
4. Jaffe ES, Lee Harris N, Stein H, Vardiman JW. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. IARC Press, 2008. 351p
5. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, Jaffe ES. The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms. *Blood* 2016;127:2375–90.
6. World Health Organization. International Classification of Diseases for Oncology. Third Edit. Geneva: 2013. 250p
7. Turner JJ, Morton LM, Linet MS, Clarke CA, Kadin ME, Vajdic CM, Monnereau A, Maynadie M, Chiu BC-H, Marcos-Gragera R, Costantini AS, Cerhan JR, et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood* 2010;116:e90-8.
8. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;
9. Sant M, Allemani C, Tereanu C, De Angelis R, Capocaccia R, Visser O, Marcos-Gragera R, Maynadié M, Simonetti A, Lutz JM, Berrino F, Hackl M, et al. Incidence of hematologic malignancies in Europe by morphologic subtype: Results of the HAEMACARE project. *Blood* 2010;116:3724–34.
10. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [25/1.
11. Bosetti C, Levi F, Ferlay J, Lucchini F, Negri E, La Vecchia C. Incidence and mortality from non-Hodgkin lymphoma in Europe: the end of an epidemic? *Int J cancer* 2008;123:1917–23.
12. Marcos-Gragera R, Pollan M, Chirlaque MD, Guma J, Sanchez MJ, Garau I. Attenuation of the epidemic increase in non-Hodgkin's lymphomas in Spain. *Ann Oncol Off J Eur Soc Med Oncol* 2010;21 Suppl 3:iii90-96.
13. Cote TR, Biggar RJ, Rosenberg PS, Devesa SS, Percy C, Yellin FJ, Lemp G, Hardy C, Geodert JJ, Blattner WA. Non-Hodgkin's lymphoma among people with AIDS: incidence, presentation and public health burden. AIDS/Cancer Study Group. *Int J cancer* 1997;73:645–50.
14. Morton LM, Sampson JN, Cerhan JR, Turner JJ, Vajdic CM, Wang SS, Smedby KE, de Sanjose S, Monnereau A, Benavente Y, Bracci PM, Chiu BCH, et al. Rationale and Design of the International Lymphoma Epidemiology Consortium (InterLymph) Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr* 2014;2014:1–14.

15. Morton LM, Slager SL, Cerhan JR, Wang SS, Vajdic CM, Skibola CF, Bracci PM, de Sanjose S, Smedby KE, Chiu BCH, Zhang Y, Mbulaiteye SM, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr* 2014;2014:130–44.
16. MacMahon B. Epidemiology of Hodgkin's disease. *Cancer Res* 1966;26:1189–201.
17. Swerdlow AJ. Epidemiology of Hodgkin's disease and non-Hodgkin's lymphoma. *Eur J Nucl Med Mol Imaging* 2003;30 Suppl 1:S3-12.
18. Chiu BC-H, Hou N. Epidemiology and etiology of non-hodgkin lymphoma. *Cancer Treat Res* 2015;165:1–25.
19. Alexander DD, Mink PJ, Adami H-O, Chang ET, Cole P, Mandel JS, Trichopoulos D. The non-Hodgkin lymphomas: a review of the epidemiologic literature. *Int J cancer* 2007;120 Suppl:1–39.
20. World Cancer Research Fund / American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR, 2007.
21. Cross AJ, Lim U. The role of dietary factors in the epidemiology of non-Hodgkin's lymphoma. *Leuk Lymphoma* 2006;47:2477–87.
22. Skibola CF. Obesity, diet and risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 2007;16:392–5.
23. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity, and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018. Available at dietandcancerreport.org.
24. Psaltopoulou T, Sergentanis TN, Ntanasis-Stathopoulos I, Tzanninis IG, Tsilimigras DI DM. Alcohol consumption and risk of hematological malignancies: A meta-analysis of prospective studies. *Int J cancer* 2018;143:486–95.
25. Han T-J, Li J-S, Luan X-T, Wang L, Xu H-Z. Dietary Fat Consumption and Non-Hodgkin's Lymphoma Risk: A Meta-analysis. *Nutr Cancer* 2017;69:221–8.
26. Psaltopoulou T, Ntanasis-Stathopoulos I, Tsilimigras DI, Tzanninis I-G, Gavriatopoulou M, Sergentanis TN. Micronutrient Intake and Risk of Hematological Malignancies in Adults: A Systematic Review and Meta-analysis of Cohort Studies. *Nutr Cancer* 2018;1–19.
27. Sergentanis TN, Psaltopoulou T, Ntanasis-Stathopoulos I, Liaskas A, Tzanninis I-G, Dimopoulos M-A. Consumption of fruits, vegetables, and risk of hematological malignancies: a systematic review and meta-analysis of prospective studies. *Leuk Lymphoma* 2018;59:434–47.
28. Chen G-C, Lv D-B, Pang Z, Liu Q-F. Fruits and vegetables consumption and risk of non-Hodgkin's lymphoma: a meta-analysis of observational studies. *Int J cancer* 2013;133:190–200.
29. Caini S, Masala G, Gnagnarella P, Ermini I, Russell-Edu W, Palli D, Gandini S. Food of animal origin and risk of non-Hodgkin lymphoma and multiple myeloma: A review of the literature and meta-analysis. *Crit Rev Oncol Hematol* 2016;100:16–24.
30. Wang J, Li X, Zhang D. Dairy Product Consumption and Risk of Non-Hodgkin Lymphoma: A Meta-Analysis. *Nutrients* 2016;8:120.
31. Bertrand KA, Giovannucci E, Rosner BA, Zhang SM, Laden F, Birmann BM. Dietary fat intake and risk of non-Hodgkin lymphoma in 2 large prospective cohorts. *Am J Clin Nutr* 2017;106:650–6.
32. McCullough ML, Teras LR, Shah R, Diver WR, Gaudet MM, Gapstur SM. Artificially and sugar-sweetened carbonated beverage consumption is not associated with risk of lymphoid neoplasms in older men and women. *J Nutr* 2014;144:2041–9.
33. Daniel CR, Sinha R, Park Y, Graubard BI, Hollenbeck AR, Morton LM, Cross AJ. Meat Intake Is Not Associated with Risk of Non-Hodgkin Lymphoma in a Large Prospective Cohort of U.S. Men and Women. *J Nutr* 2012;142:1074–80.

34. Kabat GC, Kim MY, Wactawski-Wende J, Shikany JM, Vitolins MZ, Rohan TE. Intake of antioxidant nutrients and risk of non-Hodgkin's Lymphoma in the Women's Health Initiative. *Nutr Cancer* 2012;64:245–54.
35. Rohrmann S, Linseisen J, Jakobsen MU, Overvad K, Raaschou-Nielsen O, Tjonneland A, Boutron-Ruault MC, Kaaks R, Becker N, Bergmann M, Boeing H, Khaw K-T, et al. Consumption of meat and dairy and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition. *Int J cancer* 2011;128:623–34.
36. Chang ET, Canchola AJ, Clarke CA, Lu Y, West DW, Bernstein L, Wang SS, Horn-Ross PL. Dietary phytocompounds and risk of lymphoid malignancies in the California Teachers Study cohort. *Cancer Causes Control* 2011;22:237–49.
37. Erber E, Maskarinec G, Lim U, Kolonel LN. Dietary vitamin D and risk of non-Hodgkin lymphoma: the multiethnic cohort. *Br J Nutr* 2010;103:581–4.
38. Tsai H-T, Cross AJ, Graubard BI, Oken M, Schatzkin A, Caporaso NE. Dietary factors and risk of chronic lymphocytic leukemia and small lymphocytic lymphoma: a pooled analysis of two prospective studies. *Cancer Epidemiol Biomarkers Prev* 2010;19:2680–4.
39. Ross JA, Kasum CM, Davies SM, Jacobs DR, Folsom AR, Potter JD. Diet and risk of leukemia in the Iowa Women's Health Study. *Cancer Epidemiol Biomarkers Prev* 2002;11:777–81.
40. Casabonne D, Gracia E, Espinosa A, Bustamante M, Benavente Y, Robles C, Costas L, Alonso E, Gonzalez-Barca E, Tardon A, Dierssen-Sotos T, Vazquez EG, et al. Fruit and vegetable intake and vitamin C transporter gene (SLC23A2) polymorphisms in chronic lymphocytic leukaemia. *Eur J Nutr* 2016;
41. Campagna M, Cocco P, Zucca M, Angelucci E, Gabbas A, Latte GC, Uras A, Rais M, Sanna S, Ennas MG. Risk of lymphoma subtypes and dietary habits in a Mediterranean area. *Cancer Epidemiol* 2015;39:1093–8.
42. Chang ET, Smedby KE, Zhang SM, Hjalgrim H, Melbye M, Ost A, Glimelius B, Wolk A, Adami H-O. Dietary factors and risk of non-hodgkin lymphoma in men and women. *Cancer Epidemiol Biomarkers Prev* 2005;14:512–20.
43. Purdue MP, Bassani DG, Klar NS, Sloan M, Kreiger N. Dietary factors and risk of non-Hodgkin lymphoma by histologic subtype: a case-control analysis. *Cancer Epidemiol Biomarkers Prev* 2004;13:1665–76.
44. Fritschi L, Ambrosini GL, Kliewer E V, Johnson KC. Dietary fish intake and risk of leukaemia, multiple myeloma, and non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 2004;13:532–7.
45. Charbonneau B, O'Connor HM, Wang AH, Liebow M, Thompson CA, Fredericksen ZS, Macon WR, Slager SL, Call TG, Habermann TM, Cerhan JR. Trans fatty acid intake is associated with increased risk and n3 fatty acid intake with reduced risk of non-hodgkin lymphoma. *J Nutr* 2013;143:672–81.
46. Luczynska A, Kaaks R, Rohrmann S, Becker S, Linseisen J, Buijsse B, Overvad K, Trichopoulou A, Valanou E, Barmpitsioti A, Masala G, Agnoli C, et al. Plasma 25-hydroxyvitamin D concentration and lymphoma risk: results of the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* 2013;98:827–38.
47. Mikhak B, Bracci PM, Gong Z. Intake of vitamins D and A and calcium and risk of non-Hodgkin lymphoma: San Francisco Bay Area population-based case-control study. *Nutr Cancer* 2012;64:674–84.
48. Holtan SG, O'Connor HM, Fredericksen ZS, Liebow M, Thompson CA, Macon WR, Micallef IN, Wang AH, Slager SL, Habermann TM, Call TG, Cerhan JR. Food-frequency questionnaire-based estimates of total antioxidant capacity and risk of non-Hodgkin lymphoma. *Int J cancer* 2012;131:1158–68.
49. Aschebrook-Kilfoy B, Ollberding NJ, Kolar C, Lawson TA, Smith SM, Weisenburger DD, Chiu BC-H. Meat intake and risk of non-Hodgkin lymphoma. *Cancer Causes*

Control 2012;23:1681–92.
50. Chiu BC-H, Kwon S, Evens AM, Surawicz T, Smith SM, Weisenburger DD. Dietary intake of fruit and vegetables and risk of non-Hodgkin lymphoma. *Cancer Causes Control* 2011;22:1183–95.
51. Koutros S, Zhang Y, Zhu Y, Mayne ST, Zahm SH, Holford TR, Leaderer BP, Boyle P, Zheng T. Nutrients contributing to one-carbon metabolism and risk of non-Hodgkin lymphoma subtypes. *Am J Epidemiol* 2008;167:287–94.
52. Chang ET, Balter KM, Torrang A, Smedby KE, Melbye M, Sundstrom C, Glimelius B, Adami H-O. Nutrient intake and risk of non-Hodgkin's lymphoma. *Am J Epidemiol* 2006;164:1222–32.
53. Yang L, Dong J, Jiang S, Shi W, Xu X, Huang H, You X, Liu H. Red and Processed Meat Consumption Increases Risk for Non-Hodgkin Lymphoma: A PRISMA-Compliant Meta-Analysis of Observational Studies. *Medicine (Baltimore)* 2015;94:e1729.
54. Fallahzadeh H, Cheraghi M, Amoori N, Alaf M. Red meat intake and risk of non-Hodgkin lymphoma: a meta-analysis. *Asian Pac J Cancer Prev* 2014;15:10421–5.
55. Lu D, Chen J, Jin J. Vitamin D status and risk of non-Hodgkin lymphoma: a meta-analysis. *Cancer Causes Control* 2014;25:1553–63.
56. Hu FB. Dietary pattern analysis: A new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002;13:3–9.
57. Cespedes EM, Hu FB. Dietary patterns: from nutritional epidemiologic analysis to national guidelines. *Am J Clin Nutr* 2015;101:899–900.
58. Mayne ST, Playdon MC, Rock CL. Diet, nutrition, and cancer: past, present and future. *Nat Rev Clin Oncol* 2016;13:504–15.
59. Schulze MB, Martinez-Gonzalez MA, Fung TT, Lichtenstein AH, Forouhi NG. Food based dietary patterns and chronic disease prevention. *BMJ* 2018;361:k2396.
60. Krebs-Smith SM, Pannucci TE, Subar AF, Kirkpatrick SI, Lerman JL, Tooze JA, Wilson MM, Reedy J. Update of the Healthy Eating Index: HEI-2015. *J Acad Nutr Diet* 2018;118:1591–602.
61. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH, Karanja N. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *N Engl J Med* 1997;336:1117–24.
62. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, Vassilakou T, Lipworth L, Trichopoulos D. Diet and overall survival in elderly people. *BMJ* 1995;311:1457–60.
63. Mila-Villarroel R, Bach-Faig A, Puig J, Puchal A, Farran A, Serra-Majem L, Carrasco JL. Comparison and evaluation of the reliability of indexes of adherence to the Mediterranean diet. *Public Health Nutr* 2011;14:2338–45.
64. Gleason PM, Boushey CJ, Harris JE, Zoellner J. Publishing nutrition research: a review of multivariate techniques--part 3: data reduction methods. *J Acad Nutr Diet* 2015;115:1072–82.
65. Schwingshackl L, Schwedhelm C, Galbete C, Hoffmann G. Adherence to Mediterranean Diet and Risk of Cancer: An Updated Systematic Review and Meta-Analysis. *Nutrients* 2017;9.
66. Schwingshackl L, Bogensberger B, Hoffmann G. Diet Quality as Assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: An Updated Systematic Review and Meta-Analysis of Cohort Studies. *J Acad Nutr Diet* 2018;118:74–100.e11.
67. Jayedi A, Emadi A, Shab-Bidar S. Dietary Inflammatory Index and Site-Specific Cancer Risk: A Systematic Review and Dose-Response Meta-Analysis. *Adv Nutr* 2018;9:388–403.
68. Romaguera D. Operationalising the WCRF/AICR cancer prevention recommendations using an index score: recent meta-analysis in relation to cancer

incidence and mortality. Presented at the World Cancer Congress (2018, October) Kuala Lumpur, Malaysia.
69. Grosso G, Bella F, Godos J, Sciacca S, Del Rio D, Ray S, Galvano F, Giovannucci EL. Possible role of diet in cancer: systematic review and multiple meta-analyses of dietary patterns, lifestyle factors, and cancer risk. *Nutr Rev* 2017;75:405–19.
70. Trichopoulou A, Lagiou P. Healthy traditional Mediterranean diet: an expression of culture, history, and lifestyle. *Nutr Rev* 1997;55:383–9.
71. Willett WC, Sacks F, Trichopoulou A, Drescher G, Ferro-Luzzi A, Helsing E, Trichopoulos D. Mediterranean diet pyramid: a cultural model for healthy eating. *Am J Clin Nutr* 1995;61:1402S–1406S.
72. Buckland G, Gonzalez CA, Agudo A, Vilardell M, Berenguer A, Amiano P, Ardanaz E, Arriola L, Barricarte A, Basterretxea M, Chirlaque MD, Cirera L, et al. Adherence to the Mediterranean diet and risk of coronary heart disease in the Spanish EPIC Cohort Study. *Am J Epidemiol* 2009;170:1518–29.
73. Buckland G, Travier N, Cottet V, Gonzalez CA, Lujan-Barroso L, Agudo A, Trichopoulou A, Lagiou P, Trichopoulos D, Peeters PH, May A, Bueno-de-Mesquita HB, et al. Adherence to the mediterranean diet and risk of breast cancer in the European prospective investigation into cancer and nutrition cohort study. *Int J cancer* 2013;132:2918–27.
74. Sofi F, Macchi C, Abbate R, Gensini GF, Casini A. Mediterranean diet and health status: an updated meta-analysis and a proposal for a literature-based adherence score. *Public Health Nutr* 2014;17:2769–82.
75. da Silva R, Bach-Faig A, Raido Quintana B, Buckland G, Vaz de Almeida MD, Serra-Majem L. Worldwide variation of adherence to the Mediterranean diet, in 1961-1965 and 2000-2003. *Public Health Nutr* 2009;12:1676–84.
76. Kennedy ET, Ohls J, Carlson S, Fleming K. The Healthy Eating Index: design and applications. *J Am Diet Assoc* 1995;95:1103–8.
77. Guenther PM, Reedy J, Krebs-Smith SM. Development of the Healthy Eating Index-2005. *J Am Diet Assoc* 2008;108:1896–901.
78. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HAB, Kuczynski KJ, Kahle LL, Krebs-Smith SM. Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet* 2013;113:569–80.
79. McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, Spiegelman D, Hunter DJ, Colditz GA, Willett WC. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr* 2002;76:1261–71.
80. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* 2012;142:1009–18.
81. Nutrition and Your Health: 2015-2020 Dietary Guidelines for Americans (8th edition), US Government Printing Office, Washington, DC (2015).
82. Folsom AR, Parker ED, Harnack LJ. Degree of concordance with DASH diet guidelines and incidence of hypertension and fatal cardiovascular disease. *Am J Hypertens* 2007;20:225–32.
83. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch Intern Med* 2008;168:713–20.
84. Bhupathiraju SN, Tucker KL. Coronary heart disease prevention: nutrients, foods, and dietary patterns. *Clin Chim Acta* 2011;412:1493–514.
85. Romaguera D, Vergnaud AC, Peeters PH, Van Gils CH, Chan DSM, Ferrari P, Romieu I, Jenab M, Slimani N, Clavel-Chapelon F, Fagherazzi G, Perquier F, et al. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer?

Results from the EPIC study. *Am J Clin Nutr* 2012;96:150–63.
86. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell* 2010;140:883–99.
87. Shivappa N, Steck SE, Hurley TG, Hussey JR, Hebert JR. Designing and developing a literature-derived, population-based dietary inflammatory index. *Public Health Nutr* 2014;17:1689–96.
88. Agudo A, Cayssials V, Bonet C, Tjønneland A, Overvad K, Boutron-Ruault MC, Affret A, Fagherazzi G, Katzke V, Schübel R, Trichopoulou A, Karakatsani A, et al. Inflammatory potential of the diet and risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Am J Clin Nutr* 2018;107:607–16.
89. Fowler ME, Akinyemiju TF. Meta-analysis of the association between dietary inflammatory index (DII) and cancer outcomes. *Int J cancer* 2017;141:2215–27.
90. Li D, Hao X, Li J, Wu Z, Chen S, Lin J, Li X, Dong Y, Na Z, Zhang Y, Dai H, Song Y. Dose-response relation between dietary inflammatory index and human cancer risk: evidence from 44 epidemiologic studies involving 1,082,092 participants. *Am J Clin Nutr* 2018;107:371–88.
91. Erber E, Maskarinec G, Gill JK, Park S-Y, Kolonel LN. Dietary patterns and the risk of non-Hodgkin lymphoma: the Multiethnic Cohort. *Leuk Lymphoma* 2009;50:1269–75.
92. Epstein MM, Chang ET, Zhang Y, Fung TT, Batista JL, Ambinder RF, Zheng T, Mueller NE, Birmann BM. Dietary pattern and risk of hodgkin lymphoma in a population-based case-control study. *Am J Epidemiol* 2015;182:405–16.
93. Ollberding NJ, Aschebrook-Kilfoy B, Caces DBD, Smith SM, Weisenburger DD, Chiu BC-H. Dietary patterns and the risk of non-Hodgkin lymphoma. *Public Health Nutr* 2014;17:1531–7.
94. Shivappa N, Hebert JR, Taborelli M, Montella M, Libra M, Zucchetto A, Crispo A, Grimaldi M, La Vecchia C, Serraino D, Polesel J. Dietary inflammatory index and non-Hodgkin lymphoma risk in an Italian case-control study. *Cancer Causes Control* 2017;28:791–9.
95. Shivappa N, Hébert JR, Taborelli M, Zucchetto A, Montella M, Libra M, La Vecchia C, Serraino D, Polesel J. Association between dietary inflammatory index and Hodgkin's lymphoma in an Italian case-control study. *Nutrition* 2018;53:43–8.
96. Chihara D, Ito H, Matsuda T, Shibata A, Katsumi A, Nakamura S, Tomotaka S, Morton LM, Weisenburger DD, Matsuo K. Differences in incidence and trends of haematological malignancies in Japan and the United States. *Br J Haematol* 2014;164:536–45.
97. Clarke CA, Glaser SL, Gomez SL, Wang SS, Keegan TH, Yang J, Chang ET. Lymphoid malignancies in U.S. Asians: incidence rate differences by birthplace and acculturation. *Cancer Epidemiol Biomarkers Prev* 2011;20:1064–77.
98. Evens AM, Antillón M, Aschebrook-Kilfoy B, Chiu BC-H. Racial disparities in Hodgkin's lymphoma: a comprehensive population-based analysis. *Ann Oncol* 2012;23:2128–37.
99. Castello A, Pollan M, Buijsse B, Ruiz A, Casas AM, Baena-Canada JM, Lope V, Antolin S, Ramos M, Munoz M, Lluch A, de Juan-Ferre A, et al. Spanish Mediterranean diet and other dietary patterns and breast cancer risk: case-control EpiGEICAM study. *Br J Cancer* 2014;111:1454–62.
100. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;26 Suppl 1:S6-14.
101. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data

collection. *Public Health Nutr* 2002;5:1113–24.
102. Margetts BM, Pietinen P. European Prospective Investigation into Cancer and Nutrition: validity studies on dietary assessment methods. *Int. J. Epidemiol.*1997;26 Suppl 1:S1-5.
103. Imhof A, Froehlich M, Brenner H, Boeing H, Pepys MB, Koenig W. Effect of alcohol consumption on systemic markers of inflammation. *Lancet (London, England)* 2001;357:763–7.
104. Albert MA, Glynn RJ, Ridker PM. Alcohol consumption and plasma concentration of C-reactive protein. *Circulation* 2003;107:443–7.
105. Volpato S, Pahor M, Ferrucci L, Simonsick EM, Guralnik JM, Kritchevsky SB, Fellin R, Harris TB. Relationship of alcohol intake with inflammatory markers and plasminogen activator inhibitor-1 in well-functioning older adults: the Health, Aging, and Body Composition study. *Circulation* 2004;109:607–12.
106. Buckland G, Agudo A, Lujan L, Jakszyn P, Bueno-de-Mesquita HB, Palli D, Boeing H, Carneiro F, Krogh V, Sacerdote C, Tumino R, Panico S, et al. Adherence to a Mediterranean diet and risk of gastric adenocarcinoma within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study. *Am J Clin Nutr* 2010;91:381–90.
107. Castano-Vinyals G, Aragones N, Perez-Gomez B, Martin V, Llorca J, Moreno V, Altzibar JM, Ardanaz E, de Sanjose S, Jimenez-Moleon JJ, Tardon A, Alguacil J, et al. Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. *Gac Sanit* 2015;29:308–15.
108. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, Hillmen P, Keating MJ, Montserrat E, Rai KR, Kipps TJ. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: A report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*2008;111:5446–56.
109. Garcia-Closas R, Garcia-Closas M, Kogevinas M, Malats N, Silverman D, Serra C, Tardon A, Carrato A, Castano-Vinyals G, Dosemeci M, Moore L, Rothman N, et al. Food, nutrient and heterocyclic amine intake and the risk of bladder cancer. *Eur J Cancer* 2007;43:1731–40.
110. Calvert C, Cade J, Barrett JH, Woodhouse A. Using cross-check questions to address the problem of mis-reporting of specific food groups on Food Frequency Questionnaires. UKWCS Steering Group. United Kingdom Women's Cohort Study Steering Group. *Eur J Clin Nutr* 1997;51:708–12.
111. Tablas de composición de alimentos del CESNID [Internet]. [cited 2017 Nov 9];Available from: http://www.sennutricion.org/es/2013/05/13/tablas-de-composicin-de-alimentos-del-cesnid
112. Castello A, Buijsse B, Martin M, Ruiz A, Casas AM, Baena-Canada JM, Pastor-Barriuso R, Antolin S, Ramos M, Munoz M, Lluch A, de Juan-Ferre A, et al. Evaluating the Applicability of Data-Driven Dietary Patterns to Independent Samples with a Focus on Measurement Tools for Pattern Similarity. *J Acad Nutr Diet* 2016;116:1914–1924.e6.
113. Wu S-J, Huang S-Y, Lin C-T, Lin Y-J, Chang C-J, Tien H-F. The incidence of chronic lymphocytic leukemia in Taiwan, 1986-2005: a distinct increasing trend with birth-cohort effect. *Blood* 2010;116:4430–5.
114. Monteiro CA, Cannon G, Moubarac J-C, Levy RB, Louzada MLC, Jaime PC. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutr* 2018;21:5–17.
115. Fung TT, Rimm EB, Spiegelman D, Rifai N, Tofler GH, Willett WC, Hu FB. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *Am J Clin Nutr* 2001;73:61–7.
116. Lopez-Garcia E, Schulze MB, Fung TT, Meigs JB, Rifai N, Manson JE, Hu FB. Major

dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. *Am J Clin Nutr* 2004;80:1029–35.
117. Nettleton JA, Steffen LM, Mayer-Davis EJ, Jenny NS, Jiang R, Herrington DM, Jacobs DRJ. Dietary patterns are associated with biochemical markers of inflammation and endothelial activation in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr* 2006;83:1369–79.
118. Castello A, Boldo E, Perez-Gomez B, Lope V, Altzibar JM, Martin V, Castano-Vinyals G, Guevara M, Dierssen-Sotos T, Tardon A, Moreno V, Puig-Vives M, et al. Adherence to the Western, Prudent and Mediterranean dietary patterns and breast cancer risk: MCC-Spain study. *Maturitas* 2017;103:8–15.
119. Castello A, Amiano P, Fernandez de Larrea N, Martin V, Alonso MH, Castano-Vinyals G, Perez-Gomez B, Olmedo-Requena R, Guevara M, Fernandez-Tardon G, Dierssen-Sotos T, Llorens-Ivorra C, et al. Low adherence to the western and high adherence to the mediterranean dietary patterns could prevent colorectal cancer. *Eur J Nutr* 2018;
120. Castello A, Fernandez de Larrea N, Martin V, Davila-Batista V, Boldo E, Guevara M, Moreno V, Castano-Vinyals G, Gomez-Acebo I, Fernandez-Tardon G, Peiro R, Olmedo-Requena R, et al. High adherence to the Western, Prudent, and Mediterranean dietary patterns and risk of gastric adenocarcinoma: MCC-Spain study. *Gastric Cancer* 2018;21:372–82.
121. Castello A, Boldo E, Amiano P, Castano-Vinyals G, Aragones N, Gomez-Acebo I, Peiro R, Jimenez-Moleon JJ, Alguacil J, Tardon A, Cecchini L, Lope V, et al. Mediterranean Dietary Pattern is Associated with Low Risk of Aggressive Prostate Cancer: MCC-Spain Study. *J Urol* 2018;199:430–7.
122. Bamia C, Lagiou P, Buckland G, Grioni S, Agnoli C, Taylor AJ, Dahm CC, Overvad K, Olsen A, Tjonneland A, Cottet V, Boutron-Ruault M-C, et al. Mediterranean diet and colorectal cancer risk: results from a European cohort. *Eur J Epidemiol* 2013;28:317–28.
123. Buckland G, Ros MM, Roswall N, Bueno-de-Mesquita HB, Travier N, Tjonneland A, Kiemeney LA, Sacerdote C, Tumino R, Ljungberg B, Gram IT, Weiderpass E, et al. Adherence to the Mediterranean diet and risk of bladder cancer in the EPIC cohort study. *Int J cancer* 2014;134:2504–11.
124. Molina-Montes E, Sanchez M-J, Buckland G, Bueno-de-Mesquita HBA, Weiderpass E, Amiano P, Wark PA, Kuhn T, Katzke V, Huerta JM, Ardanaz E, Quiros JR, et al. Mediterranean diet and risk of pancreatic cancer in the European Prospective Investigation into Cancer and Nutrition cohort. *Br J Cancer* 2017;116:811–20.
125. Tran H, Nourse J, Hall S, Green M, Griffiths L, Gandhi MK. Immunodeficiency-associated lymphomas. *Blood Rev* 2008;22:261–81.
126. Smedby KE, Baecklund E, Askling J. Malignant lymphomas in autoimmunity and inflammation: a review of risks, risk factors, and lymphoma characteristics. *Cancer Epidemiol Biomarkers Prev* 2006;15:2069–77.
127. Smedby KE, Askling J, Mariette X, Baecklund E. Autoimmune and inflammatory disorders and risk of malignant lymphomas--an update. *J Intern Med* 2008;264:514–27.
128. Smedby KE, Ponzoni M. The aetiology of B-cell lymphoid malignancies with a focus on chronic inflammation and infections. *J Intern Med* 2017;282:360–70.
129. Purdue MP, Lan Q, Martinez-Maza O, Oken MM, Hocking W, Huang W-Y, Baris D, Conde B, Rothman N. A prospective study of serum soluble CD30 concentration and risk of non-Hodgkin lymphoma. *Blood* 2009;114:2730–2.
130. De Roos AJ, Mirick DK, Edlefsen KL, LaCroix AZ, Kopecky KJ, Madeleine MM, Magpantay L, Martinez-Maza O. Markers of B-cell activation in relation to risk of non-Hodgkin lymphoma. *Cancer Res* 2012;72:4733–43.
131. Vermeulen R, Hosnijeh FS, Portengen L, Krogh V, Palli D, Panico S, Tumino R,

Sacredote C, Purdue M, Lan Q, Rothman N, Vineis P. Circulating soluble CD30 and future risk of lymphoma; evidence from two prospective studies in the general population. *Cancer Epidemiol biomarkers Prev* 2011;20:1925–7.

132. Purdue MP, Hofmann JN, Kemp TJ, Chaturvedi AK, Lan Q, Park J-H, Pfeiffer RM, Hildesheim A, Pinto LA, Rothman N. A prospective study of 67 serum immune and inflammation markers and risk of non-Hodgkin lymphoma. *Blood* 2013;122:951–7.
133. Purdue MP, Lan Q, Bagni R, Hocking WG, Baris D, Reding DJ, Rothman N. Prediagnostic serum levels of cytokines and other immune markers and risk of non-hodgkin lymphoma. *Cancer Res* 2011;71:4898–907.
134. Gu Y, Shore RE, Arslan AA, Koenig KL, Liu M, Ibrahim S, Lokshin AE, Zeleniuch-Jacquotte A. Circulating cytokines and risk of B-cell non-Hodgkin lymphoma: a prospective study. *Cancer Causes Control* 2010;21:1323–33.
135. Saberi Hosnijeh F, Krop EJM, Scoccianti C, Krogh V, Palli D, Panico S, Tumino R, Sacredote C, Nawroly N, Portengen L, Linseisen J, Vineis P, et al. Plasma cytokines and future risk of non-Hodgkin lymphoma (NHL): a case-control study nested in the Italian European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev* 2010;19:1577–84.
136. Conroy SM, Maskarinec G, Morimoto Y, Franke AA, Cooney R V, Wilkens LR, Goodman MT, Hernadez BY, Le Marchand L, Henderson BE, Kolonel LN. Non-hodgkin lymphoma and circulating markers of inflammation and adiposity in a nested case-control study: the multiethnic cohort. *Cancer Epidemiol Biomarkers Prev* 2013;22:337–47.
137. Berger E, Delpierre C, Hosnijeh FS, Kelly-Irving M, Portengen L, Bergdahl IA, Johansson A-S, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, et al. Association between low-grade inflammation and Breast cancer and B-cell Myeloma and Non-Hodgkin Lymphoma: findings from two prospective cohorts. *Sci Rep* 2018;8:10805.
138. Epstein MM, Rosner B, Breen EC, Batista JL, Giovannucci EL, Magpantay L, Aster JC, Rodig SJ, Bertrand KA, Laden F, Martínez-Maza O, Birmann BM. Pre-diagnosis plasma immune markers and risk of non-Hodgkin lymphoma in two prospective cohort studies. *Haematologica* 2018;haematol.2017.183236.
139. Hosnijeh FS, Heederik D, Vermeulen R. A review of the role of lymphoma markers and occupational and environmental exposures. *Vet Q* 2012;32:61–73.
140. Barbaresko J, Koch M, Schulze MB, Nothlings U. Dietary pattern analysis and biomarkers of low-grade inflammation: a systematic literature review. *Nutr Rev* 2013;71:511–27.
141. Thorburn AN, Macia L, Mackay CR. Diet, metabolites, and "western-lifestyle" inflammatory diseases. *Immunity* 2014;40:833–42.
142. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* 2009;1:6ra14.
143. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell* 2014;157:121–41.
144. Tosti V, Bertozzi B, Fontana L. Health Benefits of the Mediterranean Diet: Metabolic and Molecular Mechanisms. *J Gerontol A Biol Sci Med Sci* 2018;73:318–26.
145. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst* 2011;103:1086–92.
146. Schatzkin A, Subar AF, Moore S, Park Y, Potischman N, Thompson FE, Leitzmann M, Hollenbeck A, Morrissey KG, Kipnis V. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev* 2009;18:1026–32.
147. Hennekens C, Buring J. Epidemiology in medicine. Boston-Toronto: Company., Little Brown and, 1987.
148. Slimani N, Kaaks R, Ferrari P, Casagrande C, Clavel-Chapelon F, Lotze G, Kroke A,

Trichopoulos D, Trichopoulou A, Lauria C, Bellegotti M, Ocke MC, et al. European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. *Public Health Nutr* 2002;5:1125–45.

149. Ferrari P, Day NE, Boshuizen HC, Roddam A, Hoffmann K, Thiebaut A, Pera G, Overvad K, Lund E, Trichopoulou A, Tumino R, Gullberg B, et al. The evaluation of the diet/disease relation in the EPIC study: considerations for the calibration and the disease models. *Int J Epidemiol* 2008;37:368–78.
150. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158:14–6.
151. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, Albanes D, Gu F, Kontto J, Lassale C, Liao LM, Mannisto S, et al. Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr* 2017;105:450–65.
152. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* 1997;65:1220S–1228S; discussion 1229S–1231S.
153. Saadatian-Elahi M, Slimani N, Chajes V, Jenab M, Goudable J, Biessy C, Ferrari P, Byrnes G, Autier P, Peeters PHM, Ocke M, Bueno de Mesquita B, et al. Plasma phospholipid fatty acid profiles and their association with food intakes: results from a cross-sectional study within the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* 2009;89:331–46.
154. Waijers PMCM, Feskens EJM, Ocke MC. A critical review of predefined diet quality scores. *Br J Nutr* 2007;97:219–31.
155. International Agency for Research on Cancer. Volume 114: Consumption of red meat and processed meat. IARC Working Group. Lyon; 6–13 September, 2015. IARC Monogr Eval Carcinog Risks Hum (in press). Lyon (FR):
156. Mitra D, Shaw AK, Hutchings K. Trends in incidence of childhood cancer in Canada, 1992-2006. *Chronic Dis Inj Can* 2012;32:131–9.
157. Reedy J, Subar AF, George SM, Krebs-Smith SM. Extending Methods in Dietary Patterns Research. *Nutrients* 2018;10.
158. International Agency for Research on Cancer. Volume 96: Alcohol Consumption and Ethyl Carbamate. IARC Working Group, 2010. Monogr Eval Carcinog Risks Hum (in press).
159. Sierksma A, van der Gaag MS, Kluft C, Hendriks HFJ. Moderate alcohol consumption reduces plasma C-reactive protein and fibrinogen levels; a randomized, diet-controlled intervention study. *Eur J Clin Nutr* 2002;56:1130–6.
160. Avellone G, Di Garbo V, Campisi D, De Simone R, Raneli G, Scaglione R, Licata G. Effects of moderate Sicilian red wine consumption on inflammatory biomarkers of atherosclerosis. *Eur J Clin Nutr* 2006;60:41–7.
161. Hebert JR, Shivappa N, Wirth MD, Hussey JR, Hurley TG. Perspective: The Dietary Inflammatory Index (DII)-Lessons Learned, Improvements Made, and Future Directions. *Adv Nutr* 2019;
162. Cavicchia PP, Steck SE, Hurley TG, Hussey JR, Ma Y, Ockene IS, Hebert JR. A new dietary inflammatory index predicts interval changes in serum high-sensitivity C-reactive protein. *J Nutr* 2009;139:2365–72.
163. Castelló A, Buijsse B, Martín M, Ruiz A, Casas AM, Baena-Cañada JM, Pastor-Barriuso R, Antolín S, Ramos M, Muñoz M, Lluch A, de Juan-Ferré A, et al. Evaluating the Applicability of Data-Driven Dietary Patterns to Independent Samples with a Focus on Measurement Tools for Pattern Similarity. *J Acad Nutr Diet* 2016;116:1914–1924.e6.
164. US Departments of Agriculture and Health and Human Services . Scientific Report of the 2015 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Health and Human Services and the Secretary of Agriculture. US Government

Printing Of.
165. RARECARE. RARECARE cancer list [Internet]. [cited 2019 Jan 14];Available from: http://www.rarecare.eu/rarecancers/rarecancers.asp
166. Boffetta P, Armstrong B, Linet M, Kasten C, Cozen W, Hartge P. Consortia in cancer epidemiology: lessons from InterLymph. *Cancer Epidemiol Biomarkers Prev* 2007;16:197–9.
167. Pronk NP, Anderson LH, Crain AL, Martinson BC, O'Connor PJ, Sherwood NE, Whitebird RR. Meeting recommendations for multiple healthy lifestyle factors. Prevalence, clustering, and predictors among adolescent, adult, and senior health plan members. *Am J Prev Med* 2004;27:25–33.
168. Leech RM, McNaughton SA, Timperio A. The clustering of diet, physical activity and sedentary behavior in children and adolescents: a review. *Int J Behav Nutr Phys Act* 2014;11:4.
169. Cho ER, Shin A, Lim S-Y, Kim J. Dietary patterns and their associations with health behaviours in Korea. *Public Health Nutr* 2011;14:356–64.
170. Fewell Z, Davey Smith G, Sterne JAC. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol* 2007;166:646–55.
171. Martinez-Gonzalez MA, Corella D, Salas-Salvado J, Ros E, Covas MI, Fiol M, Warnberg J, Aros F, Ruiz-Gutierrez V, Lamuela-Raventos RM, Lapetra J, Munoz MA, et al. Cohort profile: design and methods of the PREDIMED study. *Int J Epidemiol* 2012;41:377–85.
172. Toledo E, Salas-Salvado J, Donat-Vargas C, Buil-Cosiales P, Estruch R, Ros E, Corella D, Fito M, Hu FB, Aros F, Gomez-Gracia E, Romaguera D, et al. Mediterranean Diet and Invasive Breast Cancer Risk Among Women at High Cardiovascular Risk in the PREDIMED Trial: A Randomized Clinical Trial. *JAMA Intern Med* 2015;175:1752–60.
173. Balanza R, Garcia-Lorda P, Perez-Rodrigo C, Aranceta J, Bonet MB, Salas-Salvado J. Trends in food availability determined by the Food and Agriculture Organization's food balance sheets in Mediterranean Europe in comparison with other European areas. *Public Health Nutr* 2007;10:168–76.
174. Jenkins DJ, Wolever TM, Taylor RH, Barker H, Fielden H, Baldwin JM, Bowling AC, Newman HC, Jenkins AL, Goff D V. Glycemic index of foods: a physiological basis for carbohydrate exchange. *Am J Clin Nutr* 1981;34:362–6.
175. McKenzie F, Ferrari P, Freisling H, Chajes V, Rinaldi S, de Batlle J, Dahm CC, Overvad K, Baglietto L, Dartois L, Dossus L, Lagiou P, et al. Healthy lifestyle and risk of breast cancer among postmenopausal women in the European Prospective Investigation into Cancer and Nutrition cohort study. *Int J cancer* 2015;136:2640–8.

ANNEX

Annex 1: Other related publications during the thesis period

1. Olmedo-Requena R, González-Donquiles C, Dávila-Batista V, Romaguera D, Castelló A, Molina de la Torre AJ, Amiano P, Dierssen-Sotos T, Guevara M, Fernández-Tardón G, Lozano-Lorca M, Alguacil J, Peiró R, Huerta JM, Gracia-Lavedan E, Aragonés N, Fernández-Villa T, Solans M, Gómez-Acebo I, Castaño-Vinyals G, Kogevinas M, Pollán M and Martín V. Agreement among Mediterranean Diet Pattern Adherence Indexes: MCC-Spain Study. *Nutrients* 2019. doi: 10.3390/nu11030488.

2. Roca-Barceló A, Viñas G, Pla H, Carbó A, Comas R, Izquierdo A, Pinheiro PS, Vilardell L, Solans M, Marcos-Gragera R. Mortality of women with ductal carcinoma in situ of the breast: a population-based study from the Girona province, Spain (1994–2013). *Clin Transl Oncol* 2018 Dec 10 doi: 10.1007/s12094-018-1994-1 [Epub ahead of print].

3. Solans M, Fàbrega A, Morea D, Auñon-Sanz C, Gradana I, Roncero JM, Blanco A, Kelleher N, Buch J, Saez M, Marcos-Gragera R. Population-based incidence of lymphoid neoplasms: Twenty years of epidemiological data in the Girona province, Spain. *Cancer Epidemiology* 2019; 58:8-11.

4. Solans M**,** Coenders G, Marcos-Gragera R, Gracia-Lavedan E, Benavente Y, Moreno V, Pérez-Gómez B, Amiano P, Fernández-Villa T, Guevara M, Gómez-Acebo I, Fernández-Tardón G, Vanaclocha-Espi M, Chirlaque MD, Capelo R, Barrios R, Aragonés N, Molinuevo A, Vitelli-Storelli F, Castilla J, Dierssen-Sotos T, Castaño-Vinyals G, Kogevinas M, Pollán M, Saez M. Compositional analysis of dietary paterns. *Stat Methods Med Res* 2018. doi: 10.1177/0962280218790110.

5. Fuentes-Raspall R, Solans M, Auñon-Sanz C, Saez M, Marcos-Gragera R. Incidence and survival of primary central nervous system lymphoma (PCNSL): results from the Girona cancer registry (1994–2013). *Clin Transl Oncol*. 2018; 20:1628-1630.

6. Marcos-Gragera R*, Solans M*, Galceran J, Fernández-Delgado R, Fernández-Teijeiro A, Mateos A, Quirós-Garcia J, Fuster-Camarena N, De Castro V, Sánchez MJ, Franch P, Chirlaque MD, Ardanaz E, Marcos C, Salmerón D, Peris-Bonet R; Spanish Childhood Cancer Epidemiology Working Group.

Childhood and adolescent lymphoma in Spain: incidence and survival trends over 20 years. *Clin Transl Oncol* 2018; 20:1289-1301.

7. Solans M, Osca-Gelis G, Comas R, Roncero JM, Gallardo D, Marcos-Gragera R, Saez M. Challenges in assessing the real incidence of chronic lymphocytic leukemia: 16 years of epidemiological data from the province of Girona, Spain. *Cancer Causes Control* 2018; 29(3):379-382.

8. Boldo E, Castelló A, Aragonés N, Amiano P, Pérez-Gómez B, Castaño-Vinyals G, Martín V, Guevara M, Urtiaga C, Dierssen-Sotos T, Fernández-Tardón G, Moreno V, Solans M, Peiró R, Capelo R, Gómez-Acebo I, Castilla J, Molina AJ, Castells X, Altzibar JM, Lope V, Kogevinas M, Romieu I, Pollán M on behalf of MCC-Spain researchers. Meat intake, methods and degrees of cooking and breast cancer risk in the MCC-Spain study. *Maturitas* 2018; 110:62-70.

9. Fuentes-Raspall R, Solans M, Roca-Barceló A, Vilardell L, Puigdemont M, Del Barco S, Comas R, García-Velasco A, Astudillo A, Carmona-Garcia MC, Marcos-Gragera R. Descriptive epidemiology of primary malignant and non-malignant central nervous tumors in Spain: Results from the Girona Cancer Registry (1994-2013*). Cancer Epidemiol* 2017;50(Pt A):1-8.

10. Solans M, Serra L, Osca-Gelis G, Comas R, Renart R, Vilardell L, Gallardo D, Marcos Gragera R. Incidence and survival of Hodgkin lymphoma patients in Girona (Spain) over three decades: a population-based study. *Eur J Cancer Prev* 2017;S164-S169.

Annex 2: Compositional analysis of dietary patterns

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
0(0) 1–14

DOI: 10.1177/0962280218790110
journals.sagepub.com/home/smm
SAGE

Compositional analysis of dietary patterns

M Solans,[1,2,3], G Coenders,[1,2], R Marcos-Gragera,[2,3], A Castelló,[1,4,5], E Gràcia-Lavedan,[1,6,7], Y Benavente,[8], V Moreno,[1,9,10,11], B Pérez-Gómez,[1,12], P Amiano,[1,13], T Fernández-Villa,[14], M Guevara,[1,15], I Gómez-Acebo,[1,16], G Fernández-Tardón,[1,17], M Vanaclocha-Espi,[18], MD Chirlaque,[1,19], R Capelo,[20], R Barrios,[21], N Aragonés,[1,22], A Molinuevo,[1], F Vitelli-Storelli,[14], J Castilla,[1,15], T Dierssen-Sotos,[1,16], G Castaño-Vinyals,[1,6,7,23], M Kogevinas,[1,6,7,23], M Pollán[1,4] and M Saez[1,2]

Abstract

Instead of looking at individual nutrients or foods, dietary pattern analysis has emerged as a promising approach to examine the relationship between diet and health outcomes. Despite dietary patterns being compositional (i.e. usually a higher intake of some foods implies that less of other foods are being consumed), compositional data analysis has not yet been applied in this setting. We describe three compositional data analysis approaches (compositional principal component analysis, balances and principal balances) that enable the extraction of dietary patterns by using control subjects from the Spanish multicase-control (MCC-Spain) study. In particular, principal balances overcome the limitations of purely data-driven or investigator-driven methods and present dietary patterns as trade-offs between eating more of some foods and less of others.

Keywords

Compositional data analysis, dietary patterns, epidemiology, principal balances, MCC-Spain

[1]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain
[2]Research Group on Statistics, Econometrics and Health (GRECS), Universitat de Girona, Girona, Spain
[3]Epidemiology Unit and Girona Cancer Registry, Oncology Coordination Plan, Department of Health, Autonomous Government of Catalonia, Catalan Institute of Oncology, Girona, Spain
[4]Cancer Epidemiology Unit, National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain
[5]Faculty of Medicine, University of Alcalá, Alcalá de Henares, Madrid, Spain
[6]ISGlobal, Barcelona, Spain
[7]ISGlobal, Universitat Pompeu Fabra (UPF), Barcelona, Spain
[8]Unit of molecular and genetic epidemiology in infections and cancer, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain
[9]Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO). Hospitalet de Llobregat, Barcelona, Spain
[10]Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL). Hospitalet de Llobregat, Barcelona, Spain.
[11]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain
[12]Cardiovascular and Metabolic Diseases Epidemiology Unit, National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain
[13]Public Health Division of Gipuzkoa, BioDonostia Research Institute, Health Department, Basque Country, San Sebastian, Spain
[14]Instituto de Biomedicina, Universidad de León, León, Spain
[15]Instituto de Salud Pública de Navarra, IdiSNA, Pamplona, Spain
[16]Universidad de Cantabria – IDIVAL, Santander, Spain
[17]IUOPA, Universidad de Oviedo, Asturias, Spain
[18]Cancer and Public Health Area, FISABIO – Public Health, Valencia, Spain
[19]Department of Epidemiology, Regional Health Authority, IMIB-Arrixaca, Murcia University, Murcia, Spain
[20]Centro de Investigación en Recursos Naturales, Salud y medio Ambiente (RENSMA), Universidad de Huelva, Huelva, Spain
[21]Departamento de Medicina Preventiva y Salud Pública, Facultad de Medicina, Universidad de Granada, Granada, Spain
[22]Epidemiology Section, Public Health Division, Department of Health of Madrid, Madrid, Spain
[23]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

Corresponding author:
Marc Saez, Research Group on Statistics, Econometrics and Health (GRECS) and CIBER of Epidemiology and Public Health (CIBERESP) University of Girona Carrer de la Universitat de Girona 10, Campus de Montilivi, Girona 17003, Spain.
Email: marc.saez@udg.edu

1 Introduction

In nutritional epidemiological studies, there is keen interest in identifying specific dietary components that may be related to particular health outcomes. Traditionally, research has focused on single dietary factors (i.e. nutrients, foods or food groups), even though individuals do not consume them in isolation. Thus, in the recent years most studies have shifted to dietary pattern analysis, which better captures overall dietary exposure and allows the cumulative and interactive effects between dietary factors to be evaluated.[1,2] The foremost methods for extracting dietary patterns from a given population are *a priori* and *a posteriori* approaches.[3] The former are investigator-driven or index-based analyses which use a numerical scoring system defined on the basis of previous scientific evidence. Thus, indexes may differ in design, structure, and interpretation of dietary guidance (e.g. multiple indexes describe adherence to a Mediterranean diet, using different food groups, weightings and cut-offs for recommended intakes[4]), but once there is agreement on which index to use, it eases comparability across populations. The later are data-driven methods that use principal component (PC) (or factor) analysis, cluster analysis, and related techniques, to derive dietary patterns. These patterns are more representative of the eating habits of the study population and, although their applicability to a different setting has been a major concern,[1] recent evidence has proven that, under certain conditions, they may be used in different populations.[5,6]

Usually, a higher intake of some foods implies that less of other foods are being consumed. In dietary interventions that advocate an increase or decrease of particular foods or nutrients, unless total caloric intake is modified, changes in one dietary component are accompanied by compensatory changes in others. The many food pyramids which have been built represent nothing more than ideal relative amounts of food groups within a total intake. Compositional data analysis (CoDA) is a standard family of statistical methods for analyzing the relative importance of magnitudes, and holds great potential in the context of dietary patterns. Within this family of CoDA methods, in more precise terms we refer to the so-called CoDA log-ratio approach.

Although CoDA is a well-established statistical methodology in many scientific fields (e.g. geology, hydrology, or ecology),[7] it has only recently been used in health research. Health-related time-use research constitutes the most frequent application.[8–19] However, relative information also lies at the core of the research interest in nutrition,[20–23] cause-specific mortality,[24] genomics[25,26] and microbiome.[27–29] To our knowledge, no study has yet reported its application in the context of dietary patterns.

The aim of this study was to apply and compare three CoDA approaches (compositional PC analysis, balances and principal balances) that enable dietary patterns to be extracted and later used as health outcome predictors. The methods chosen ranged from more data-driven approaches to more investigator-driven ones. We illustrate the methods with data from the Spanish multicase-control (MMC-Spain) study. For this purpose, we selected a subset of food groups, which are typically used to describe adherence to a Mediterranean diet (MD).

2 Methods

2.1 Compositional data analysis basics

The use of CoDA started with Aitchison's seminal work[30,31] on chemical and geological compositions in which data are expressed as parts of a whole, commonly with a fixed sum.[32] The term compositional analysis[33] was later coined to stress the fact that what is ultimately compositional is not the data, which may not have a fixed sum[8–19] and may not even constitute parts of the same whole or of any whole at all, but the analysis and research objectives which are expressed in terms of relative importance of magnitudes.[34] In dietary research, this flexibility makes it possible to combine nutrients and food groups in the same analysis. Trichopoulou's MD index,[35] which considers both food groups and fatty acids, is a good example. In the last three decades, CoDA has provided a ready-to-use toolbox including software such as the R libraries SpiecEasi, compositions, zCompositions, propr and robCompositions,[29,36–39] the stand-alone programs SparCC and CoDaPack,[26,40] and accessible handbooks.[7,36,41]

Let the composition $\mathbf{x}$ be a positive vector in a D-dimensional real space:

$$\boldsymbol{x} = (x_1,\ x_2, \ldots, \ x_D) \in R_+^D, \quad \text{with} \quad x_j > 0 \quad \text{for all } j = 1, 2, \ldots, D \tag{1}$$

where D is the number of parts, in our case, food groups or nutrients. In order to focus on the relative importance of the parts, the *closure* of $\mathbf{x}$ to a constant unit sum is common practice.

$$\begin{aligned} z = C(x) &= \left(\frac{x_1}{S},\ \frac{x_2}{S}, \ldots, \ \frac{x_D}{S}\right) = (z_1,\ z_2, \ldots, \ z_D) \\ &\text{with } z_j > 0 \quad \text{for all } j = 1, 2, \ldots, D; \qquad \sum_{j=1}^{D} z_j = 1;\ S = \sum_{j=1}^{D} x_j \end{aligned} \tag{2}$$

In our case, each subject would have a composition of each of the D food groups as proportions of total energy intake, total grams or portions per day or week, or whichever measurement units the data are expressed in. However, closure is by no means required. Regardless of whether closure is performed or not, the relative information carried out by the D parts should remain the same, ensuring the so-called *compositional equivalence* property.[33]

$\mathbf{z}$ resides in an R_+^{D-1} subspace which is constrained by positivity and a fixed sum, called the *simplex*, with different operations, angles and distances from the real space. For this reason, most statistical workhorses such as correlation, variance, and Euclidean distance are to a lesser or greater extent meaningless when applied to $\mathbf{z}$. This has implications when studying dietary patterns with any correlation-based method, such as PC analysis.[42] Finally, when it comes to statistical modelling, distributional assumptions of classical models are violated on $\mathbf{z}$,[7,43] since constraints in $\mathbf{z}$ make it impossible to use unbounded probability distributions such as the normal distribution.

2.2 Transformations, association, and variance

The most common CoDA approach is to express an original compositional vector of D parts into logarithms of ratios among parts or of ratios among geometric means of parts.[31,44] There are six main arguments for log-ratios. First, log-ratios are unbounded and, once they have been computed, the normal distribution and other unbounded distributions can be used. Second, standard statistical analyses based on Euclidean geometry in the real space are appropriate. Third, log-ratios are compositionally equivalent, as they yield the same result regardless of whether they are computed from $\mathbf{x}$ or $\mathbf{z}$. Fourth, log-ratios form the basis for defining association, distance and variance in a geometrically meaningful way. Fifth, log-ratios treat the numerator and denominator symmetrically. Sixth, and most important for the purposes of this article, logarithms, ratios and geometric means constitute a natural way of distilling the information about the relative importance of food groups and nutrients within the dietary patterns.

Log-ratios may be computed between each part and the geometric mean of all, in the so-called *centred-log ratios*

$$\ln\left(\frac{z_j}{\sqrt[D]{z_1 z_2 \dots z_D}}\right) = \ln\left(\frac{x_j}{\sqrt[D]{x_1 x_2 \dots x_D}}\right) \quad \text{with} \quad j = 1, 2, \dots, D \tag{3}$$

A higher centred log-ratio for a given subject on food group j means a higher relative importance of that food group within total intake. The sum of all centred log-ratios for a given subject is zero. Unlike the simple log transform which is commonly used in dietary research, the centred log-ratio of a given food group can only increase if at least some other decreases.

Total variance in a compositional data set is expressed by the sum of variances of all centred log-ratios

$$\sum_{j=1}^{D} Var\left(\ln\left(\frac{z_j}{\sqrt[D]{z_1 z_2 \dots z_D}}\right)\right) \tag{4}$$

Proportionality between pairs of food groups is a valid alternative to correlation.[45] The log-ratios between all $D(D\text{-}1)$ possible pairs of parts and their variances are computed for this purpose.

$$\mathrm{Var}\left(\ln\left(z_j/z_k\right)\right) = \mathrm{Var}\left(\ln\left(z_k/z_j\right)\right) \quad \text{with} \quad j, k = 1, \dots, D; \quad j \neq k \tag{5}$$

These variances can be arranged in a symmetric matrix with parts (i.e. food groups) defining both D rows and D columns, with the same layout as a correlation matrix. It is the so-called variation matrix.[7,31] It can be shown that the sum of elements in the variation matrix is $2D$ times the total variance.[4] More advanced proportionality measures are available.[38]

$\mathrm{Var}(\ln(z_j/z_k))$ is zero when z_j and z_k behave perfectly proportionally (e.g. individuals eating twice of one food group also eat twice of the other), which corresponds to perfect positive association. The further $\mathrm{Var}(\ln(z_j/z_k))$ is from zero, the lower the association. There is no clearly defined threshold representing lack of association, so that values in the matrix must be assessed comparatively.

A relevant issue in CoDA is the so-called *subcompositional coherence* principle.[7] In dietary pattern terms, this concerns the decision on which food groups and nutrients to include in the analysis. Of course, including or excluding a food group does influence the results. However, the results obtained with a set of

food groups or with a smaller subset of the former must be mutually coherent. The log-ratio methods in CoDA ensure that:

- Distances between subjects using the full set of food groups are equal or larger than when using a subset.
- Log-ratios and log-ratio variances involving pairs of food groups which are both in the full set and in the subset, are invariant.
- Geometrically speaking, subcompositions constitute an orthogonal projection of the whole composition.

Subcompositional coherence makes it possible to exclude from the analysis food groups which are not relevant to adherence to a particular dietary guidance.

2.3 Zero replacement

As it is well known, computing log-ratios implies that **x** and **z** may contain no zero values in any food group intake. Treatment of zeros in CoDA depends on the assumed reason for their occurrence, which is deemed more important than the sheer existence of zeros in itself.[46]

On the one hand, there are *absolute, essential or structural zeros*, which represent values that can only be zero given certain characteristics of the individuals (e.g. meat or fish intake in vegetarians). The presence of structural zeros may lead to different covariance structure of the variables of interest, and usually indicates that the choice of parts to be analysed is not meaningful to a certain subpopulation. Thus, data with absolute zeros should be considered as distinct subpopulations and should either be excluded (e.g. by analysing only non-vegetarians) or analysed separately (e.g. by using other dietary scores that better apply to vegetarians).

On the other hand, the so-called *rounded zeros, trace zeros, or zeros below detection limit* constitute parts which are believed to be present, but are not observed due to randomness or limitations of measurement (e.g. a retrospective food frequency questionnaire expressed in weekly portions may fail to record food groups which are consumed less frequently). They are, thus, analogous to missing data with the added information that they are below the detection limit (e.g. the gram equivalent of one portion per week). They can thus be imputed by means of the EM algorithm if modified in such a way that no imputed value is allowed to be above the detection limit.[47]

2.4 Extracting compositional information for its use as a predictor in statistical models

The D centred log-ratios (equation (3)) play an important role in distance-based statistical methods such as PC and cluster analysis, but they are not practical as predictors in statistical models for a number of reasons. The fact that they are perfectly collinear is their most often considered disadvantage, although computational solutions do exist.[48] A more serious drawback is that each centred log-ratio, and hence each regression coefficient, is related to one particular component, which is not particularly useful when the interest of the researcher lies in dietary patterns as a whole.

Other forms of log-ratios, also called coordinates, are required, which can be interpreted in terms of dietary patterns and on which both geometrical operations and statistical models can be applied in a standard manner in a whole real space matching the (D-1)-dimensionality of the simplex.[7] This approach is referred to as *working on coordinates* in the CoDA literature.[49]

Egozcue et al.[44] establish several desirable properties that a set of log-ratios, and hence coordinates, must have in order to be used as variables in further statistical analyses. The most general expression of a log-ratio includes r parts in the numerator and s parts in the denominator, with possibly different exponents in the numerator Ψ_{nj} and in the denominator Ψ_{dj}

$$\ln \frac{\left(x_{n1}^{\psi_{n1}} \dots x_{nr}^{\psi_{nr}}\right)}{\left(x_{d1}^{\psi_{d1}} \dots x_{ds}^{\psi_{ds}}\right)} \tag{6}$$

The coordinates fulfilling all desirable properties are the so-called *isometric log-ratios*, or *isometric log-ratio coordinates*, and have the following requirements:

- They must define an orthogonal (D-1)-dimensional basis in the simplex.
- The sum of exponents in the numerator of the log-ratio must equal the sum of exponents in the denominator.

- The sum of all squared exponents must be one.

It can be proven that D-1 isometric log-ratios capture all information in the compositional data set.[44] The total variance of the D-1 isometric log-ratios equals the total variance of the D centred log-ratios (equation (4)). Either the full set of D-1 isometric log-ratios or a subset may be used both as dependent and as explanatory variables in any standard statistical model. Using a subset of these D-1 log-ratios is tantamount to an orthogonal projection into a lower-variance subspace. Three different approaches for computing either D-1 or a smaller number of isometric log-ratios are presented below.

2.5 Compositional PC coordinates

Aitchison[42] extended the well-known data-driven *PC analysis* procedure to the compositional case. The extension boils down to submitting the D centred log-ratios (equation (3)) to an otherwise standard PC analysis of the covariance matrix. D-1 PC scores with decreasing variance are extracted, from here on called *PC coordinates*. The PC coordinates are actually log-ratios (equation (6)) in which positive component loadings are the Ψ_{nj} unequal exponents of parts in the numerator and negative loadings are the Ψ_{dj} unequal exponents of parts in the denominator. The D-1 PC coordinates can be proven to fulfil all conditions for being isometric log-ratio coordinates. Either all D-1 PC coordinates or, more commonly, the first few of them explaining most of the variance, can thus be used as variables in further statistical analyses.

2.6 Balance coordinates

Isometric log-ratios can also be investigator-driven, on the basis of the investigator's research questions. As a general guideline to find D-1 investigator-driven isometric log-ratio coordinates Egozcue and Pawlowsky-Glahn[50] propose *balance coordinates*. Balance coordinates can be easily formed from a *sequential binary partition* (SBP) of parts. To create the first balance coordinate, the complete composition $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ is partitioned into two groups of parts: one for the numerator and the other for the denominator. In the following step, one of the two groups is further split into two new groups to create the second balance coordinate. In step k when the y_k balance is created, a group containing r_k+s_k parts is split into two: the r_k parts $(x_{n1},\ldots, x_{nr})$ in the first group are placed in the numerator, and the s_k parts $(x_{d1},\ldots,x_{ds})$ in the second group appear in the denominator. The balance coordinate obtained is a normalised log-ratio of the geometric means of each group of parts[44]

$$y_k = \sqrt{\frac{r_k s_k}{r_k + s_k}} \ln \frac{\sqrt[r_k]{(x_{n1} \ldots x_{nr})}}{\sqrt[s_k]{(x_{d1} \ldots x_{ds})}} = \ln \frac{(x_{n1} \ldots x_{nr})^{\sqrt{\frac{s_k}{r_k}/\frac{r_k}{r_k+s_k}}}}{(x_{d1} \ldots x_{ds})^{\sqrt{\frac{r_k}{s_k}/\frac{r_k}{r_k+s_k}}}}, \quad \text{with} \quad k = 1, \ldots, D-1 \tag{7}$$

The corresponding expression (6) of the kth balance takes equal values $\psi_{nj} = \sqrt{\frac{s_k}{r_k(r_k+s_k)}}$ for all parts in the numerator, equal values, $\psi_{dj} = \sqrt{\frac{r_k}{s_k(r_k+s_k)}}$ for parts appearing in the denominator and $\psi = 0$ for parts appearing nowhere. Positive balance coordinates show a higher relative weight of parts in the numerator, and negative values show the opposite. Normally, all D-1 balance coordinates are kept for use as variables in further statistical analyses.

Unlike hierarchical cluster analysis, SBPs and hence balance coordinates are not driven by the data but can be tailored to the research questions of interest. For this purpose, SBPs may be constructed according to conceptual similarity of parts, to theoretically meaningful comparisons of numerator and denominator parts, or to trade-offs between numerator and denominator parts which extant knowledge expects to affect a health outcome. The total variance in the D-1 balance coordinates can thus be partitioned into the variance related to the research questions which have driven the construction of the SBP.

Balance coordinates have a visualization tool called the *CoDa-dendrogram*,[51] also referred to as the balance dendrogram. It is a depiction of the SBP as a tree diagram. Each balance coordinate is represented on a horizontal axis between the two groups of parts which are divided at the corresponding SBP step. The vertical bar going up from each one of these axes represents the variance of that specific coordinate. The contact point is the coordinate mean, closer to the right set of parts if these parts are relatively more abundant, closer to the left set of parts if this

Annex

set of parts is relatively more abundant, or just in the middle if the balance coordinate mean is zero. Box plots may be added to represent the balance coordinate medians and quartiles.

2.7 Principal balances

PC coordinates are a very efficient tool to compute isometric log-ratio coordinates. The fact that the first few PC coordinates explain most of the variance makes them especially fit to summarize the composition into few variables for further statistical analyses. However, the PC coordinates obtained can be difficult to interpret as they generally involve all the parts of the composition with irregular Ψ exponents.[52] Being data driven, such coefficients would be recomputed each time the analysis was rerun on a different data set, thus making comparative research less practical.

On the other hand, balance coordinates compare readily identifiable groups of parts with equal exponents in the numerator and the denominator (actually, the geometric means of numerator and denominator parts) but they require the investigator to provide a SBP. It may prove difficult to provide a theory-driven SBP when D is large, or there may also be more than one SBP candidate. At best, selection of the SBP will always remain subjective to some extent. Besides, there is no guarantee that a small number of balance coordinates account for a large proportion of total variance.

The possibility of developing an intermediate approach sharing the best properties of PC coordinates and balance coordinates holds promise. The so-called *principal balances* first suggested by Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado,[53] are data-driven balance coordinates, in which a large proportion of variance concentrates on a few coordinates, while comparing readily identifiable groups of parts with easy-to-interpret equal Ψ exponents in the numerator and the denominator. These exponents can be easily kept for replication on other data sets. The first principal balance is defined as the balance coordinate which maximizes explained variance. Subsequent principal balances, being orthogonal to the preceding ones, also maximize the explained remaining variance. Computing principal balances exactly fulfilling this definition requires an exhaustive search along all possible SBPs. A recommended heuristic method is to use the variation matrix among parts as if it was a squared Euclidean distance matrix, and cluster parts based on this matrix with Ward's clustering algorithm.[52] The resulting classification tree diagram provides an SBP with balance coordinates which are close to being principal balances, from which, a CoDa-dendrogram can be represented. All D-1 principal balances, or, alternatively, those with the highest variance, can be used as variables in subsequent analyses. Alternative computationally intensive methods are described elsewhere.[52]

3 Application example

3.1 Study population and data preprocessing

The example uses data from the MCC-Spain study, a multicentric case-control study launched to evaluate the influence of environmental exposures and their interaction with genetic factors in four common tumors in Spain. Additional information regarding the study design is provided elsewhere.[54] In brief, between September 2008 and December 2013, subjects aged 20–85 with a histologically confirmed newly diagnosed cancer were recruited in 23 Spanish hospitals from 12 Spanish provinces. Simultaneously, population-based controls frequency-matched to cases, by age, sex and region were randomly selected from primary care centers within hospitals' catchment areas. For the current analysis, only control population was used. All participants signed an informed consent. Approval for the study was obtained from the ethical review boards of all recruiting centers.

Subjects were provided a semi-quantitative Food Frequency Questionnaire (FFQ), which was a modified version from a previously validated instrument in Spain to include regional products. It included 140 food items, and assessed usual dietary intake during the previous year. A subset of dietary components (in g/day intake) was selected for this illustrative example, based on a common pattern such as the MD as conceptualized by Trichopoulou et al.[35] Alcoholic beverages were excluded from our analyses, as they are known risk factors for several chronic diseases (e.g. cardiovascular conditions, cancer). Thus, for the current illustration, the following items were used:

x_1 Vegetables
x_2 Fruits and nuts (fruit)
x_3 Legumes
x_4 Fish and seafood (seafood)

x_5 Cereals
x_6 Meat
x_7 Dairy
x_8 Monounsaturated fats
x_9 Saturated fats

We assumed that respondents reporting no consumption of any type of meat, fish, and seafood were vegetarians, and we treated them as absolute zeros by removing them. We then replaced trace zeros. Zero percentages (3.59% overall; vegetables 0.44%, fruit 0.99%, legumes 27.16%, seafood 0.96%, cereals 0.91%, meat 0.16%, dairy 1.73%, monounsaturated fats 0%, and saturated fats 0%) were acceptable for replacement with the modified EM algorithm.[47] To check for the presence of multivariate outliers in the coordinate vector, squared Mahalanobis distances to the centre can be used.[55] After removing cases above the 99.9 percentile of the χ^2 distribution with eight degrees of freedom (167 cases), a sample size of 3471 individuals was obtained.

Table 1 shows the variation matrix and centred log-ratio variances of the nine dietary components. The variation matrix led to the observation that both fat types were those components which behaved most proportionally. This means that most individuals in our sample had either a high or low intake of both fat types and thus the comparison of the fatty acid profile may contribute little to defining a useful dietary pattern. Similarly, cereals and meat behaved quite proportionally, in spite of the fact that the former represented MD and the latter a Western-like pattern. By contrast, the highest variances corresponded to legumes, whose consumption tended to move away from that of any other component. This is also depicted by legumes having the highest centred log-ratio variance. The second and third highest variances corresponded to dairy products and fruit.

3.2 PC coordinates

Table 2 shows the PC loadings and percentages of explained variance by PC coordinates. Together, three PCs (1, 2 and 3) accounted for 79.3% of the total variance of the *D* centered log-ratios (equation (4)) and could be used as a fair summary of diet composition. The first PC coordinate basically reflected the comparison between legumes and the rest of parts, with irregular coefficients. The second basically compared dairy with seafood, vegetables, and fruit. The third PC coordinate balanced meat with dairy and fruit.

3.3 Balance coordinates

Figure 1 represents the CoDa-dendrogram corresponding to an investigator-driven SBP, in this case an adaptation of Trichopoulou's score of adherence to a Mediterranean dietary pattern.[35] At the top of the SBP as the first partition in the dendrogram, y_1 separates food and nutrient groups presumed to fit a MD (x_1 to x_5 and x_8) in the

Table 1. Variation matrix and centred log-ratio variances of the nine dietary components.

Variation matrix **(5)**

	Vegetables	Fruit	Legumes	Seafood	Cereals	Meat	Dairy	Monoun-saturated fats	Saturated fats
Vegetables	0.000	0.633	2.808	0.538	0.597	0.645	0.939	0.476	0.494
Fruit	0.633	0.000	3.201	0.720	0.714	0.897	0.995	0.669	0.677
Legumes	2.808	3.201	0.000	3.022	3.105	3.052	3.509	3.005	2.906
Seafood	0.538	0.720	3.022	0.000	0.521	0.508	0.942	0.434	0.423
Cereals	0.597	0.714	3.105	0.521	0.000	0.444	0.781	0.324	0.265
Meat	0.645	0.897	3.052	0.508	0.444	0.000	0.883	0.331	0.244
Dairy	0.939	0.995	3.509	0.942	0.781	0.883	0.000	0.725	0.586
Monounsaturated fats	0.476	0.669	3.005	0.434	0.324	0.331	0.725	0.000	0.099
Saturated fats	0.494	0.677	2.906	0.423	0.265	0.244	0.586	0.099	0.000
Centred log-ratio variances **(E4)**									
	0.062	0.096	0.487	0.062	0.053	0.059	0.117	0.036	0.027

Table 2. PC loadings (Ψ exponents) and percentages of explained variances by each PC coordinate.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Vegetables	0.053	−0.320	−0.158	0.632	0.584	−0.032	0.123	−0.037
Fruit	0.109	−0.428	−0.654	−0.456	−0.113	0.212	0.025	−0.017
Legumes	−0.939	0.066	−0.002	−0.039	−0.025	−0.009	−0.005	0.013
Seafood	0.104	−0.262	0.138	0.424	−0.761	−0.165	−0.067	−0.017
Cereals	0.136	0.006	0.187	−0.357	0.125	−0.745	0.362	0.079
Meat	0.117	0.029	0.461	−0.111	−0.006	0.604	0.514	0.145
Dairy	0.170	0.791	−0.424	0.182	−0.091	0.021	0.069	0.088
Monounsaturated fats	0.130	−0.003	0.225	−0.125	0.174	0.046	−0.682	0.555
Saturated fats	0.119	0.121	0.227	−0.149	0.112	0.069	−0.338	−0.809
Variance (%)	55.185	12.216	11.858	6.429	5.566	4.533	3.379	0.835

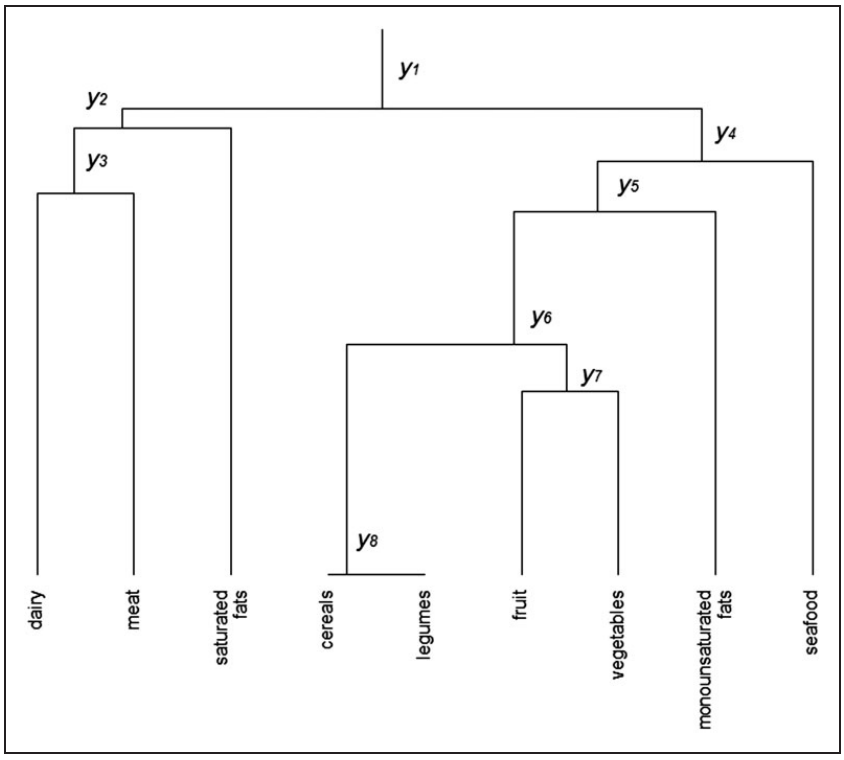


Figure 1. CoDa-dendrogram corresponding to the investigator-driven balance coordinates. Boxplots omitted for simplicity.

numerator and those not related to a MD(x_6, x_7 and x_9) in the denominator. Thus, the y_1 balance coordinate is a score of adherence to a Mediterranean dietary pattern. If the population is heterogeneous with regard to adherence, this coordinate should contribute a substantial part of total variance. In addition, below that first balance coordinate (y_1), both conglomerates of food groups are further subdivided sequentially. These subdivisions would address further research questions chosen by the investigator based on knowledge about the particular health outcome of interest. In the example presented here, they would concern the importance of the relative intake of meat and dairy (y_3) for a health outcome, the importance of the relative intake of legumes and cereals (y_8), and so on.

As PC coordinates do, balance coordinates have an implicit loading matrix following the Ψ exponents (equation (7)). Table 3 presents the Ψ exponents and the percentage of explained variance by each balance

Table 3. Investigator-driven Ψ exponents and percentages of explained variance by each investigator-driven balance.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
Vegetables	0.236	0.000	0.000	−0.183	−0.224	0.500	0.707	0.000
Fruit	0.236	0.000	0.000	−0.183	−0.224	0.500	−0.707	0.000
Legumes	0.236	0.000	0.000	−0.183	−0.224	−0.500	0.000	0.707
Seafood	0.236	0.000	0.000	0.913	0.000	0.000	0.000	0.000
Cereals	0.236	0.000	0.000	−0.183	−0.224	−0.500	0.000	−0.707
Meat	−0.471	−0.408	0.707	0.000	0.000	0.000	0.000	0.000
Dairy	−0.471	−0.408	−0.707	0.000	0.000	0.000	0.000	0.000
Monounsaturated fats	0.236	0.000	0.000	−0.183	0.894	0.000	0.000	0.000
Saturated fats	−0.471	0.816	0.000	0.000	0.000	0.000	0.000	0.000
Variance (%)	11.731	2.837	9.666	7.764	7.482	19.607	6.925	33.988

Table 4. Principal balance Ψ exponents and percentages of explained variance by each principal balance.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
Vegetables	−0.118	−0.134	0.436	−0.408	0.707	0.000	0.000	0.000
Fruit	−0.118	−0.134	0.436	0.816	0.000	0.000	0.000	0.000
Legumes	0.943	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Seafood	−0.118	−0.134	0.436	−0.408	−0.707	0.000	0.000	0.000
Cereals	−0.118	−0.134	−0.327	0.000	0.000	0.866	0.000	0.000
Meat	−0.118	−0.134	−0.327	0.000	0.000	−0.289	0.816	0.000
Dairy	−0.118	0.935	0.000	0.000	0.000	0.000	0.000	0.000
Monounsaturated fats	−0.118	−0.134	−0.327	0.000	0.000	−0.289	−0.408	0.707
Saturated fats	−0.118	−0.134	−0.327	0.000	0.000	−0.289	−0.408	−0.707
Variance (%)	54.838	11.849	10.175	7.908	5.887	4.421	3.835	1.087

coordinate, which are further shown in Figure 1. The first balance coordinate (y_1) representing adherence to a MD explained only a small portion of the variance. This shows that most heterogeneity among eating patterns lies elsewhere, mainly in the balance coordinate opposing the two main types of grains (y_8), and in the balance coordinate between grains and the combination of fruit and vegetables (y_6), as shown by their higher percentages of variance and by the longer vertical segments going up from the balance coordinates in the CoDa-dendrogram in Figure 1. Overall, cereals were consumed more than legumes, as shown by the vertical bar representing y_8 closer to the cereal side. Along similar lines, fruit and vegetables were more prevalent than grains, as shown by the vertical bar representing y_6 closer to the fruit and vegetable side.

3.4 Principal balances

The Ψ exponent matrix and the percentages of explained variance are shown in Table 4. The first and second principal balances resembled the first and second PC coordinates. In both cases they were dominated by the ratio of legumes and dairy over most or all of the remaining food groups. The third principal balance is particularly interesting as it compared vegetables, fruit and seafood with cereals, meat and fat. The comparison between monounsaturated and saturated fat had virtually no variance. Subjects were either eating more of both or less of both in nearly proportional terms (proportionality between these two parts corresponded to the lowest entry in the variation matrix). Together, the three first principal balances accounted for 76.9% of the total variance of the D centered log-ratios and could be used as a fair summary of diet composition. The lengths of the vertical bars above each principal balance in the dendrogram in Figure 2 also show their variance.

3.5 Comparison between the compositional dietary patterns

The first investigator-driven balance coordinate had a high multiple correlation with the set of three first principal balances at 0.933. This shows that the first three principal balances contained virtually all the information in the

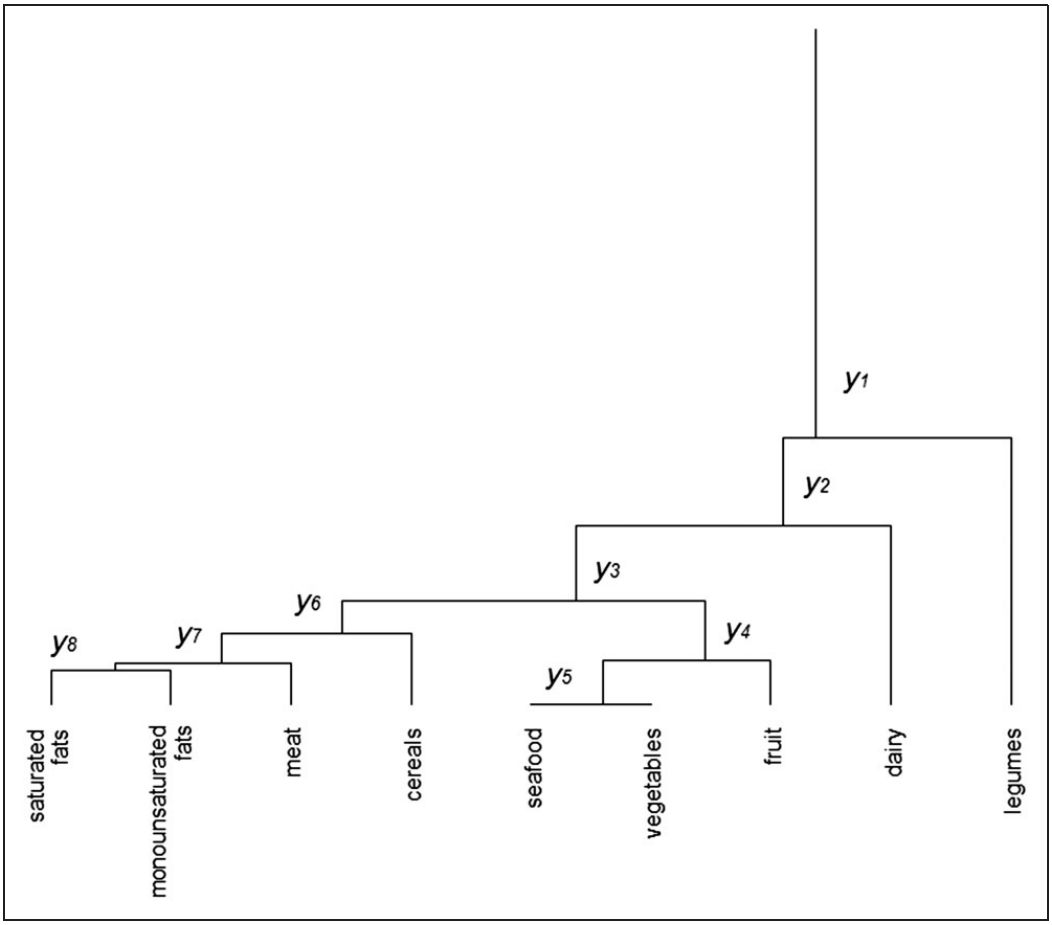


Figure 2. CoDa-dendrogram corresponding to the data-driven principal balances. Boxplots omitted for simplicity.

investigator-driven pattern, but not the other way around because correlations between the investigator-driven pattern and each of the first three principal balances were relatively low at 0.601, −0.578, and 0.577, respectively. The fact that correlations were, at best, moderate is in accordance with the fact that the first investigator-driven balance coordinate only accounted for 11.7% of the variance in dietary composition.

In the same vein, the multiple correlation between first investigator-driven balance coordinate and the first three PC coordinates was 0.941. It can thus be argued that PC coordinates and principal balances perform the job of summarizing the information in diet composition almost equally well. Principal balances would be preferable on the grounds that they are easily interpretable and lend themselves more readily to replication and comparison. In fact, the correlations between the first three PC coordinates and their corresponding principal balances were −0.999, 0.868, and −0.706, respectively (it must be noted that the negative signs have no particular implication. All correlations may be turned into positive by reversing the numerator and denominator of one of the coordinates which are being correlated).

4 Comments

This article is the first to compare several CoDA methods to extract compositional information in dietary patterns, in a manner that is appropriate for their later use as predictors of health outcomes. Predictors can be, alternatively, the first balance coordinate, all balance coordinates, the first few PC coordinates, or the first few principal balances. The model and estimation method will be dictated only by the characteristics of the dependent variable and the research design. For instance, natural choices can be a linear model for a continuous health outcome,[56] a probit or logit model for an ordered or unordered categorical health outcome,[57] and a Cox regression

for survival time. Predictors are introduced in a standard manner, the model of choice is estimated with standard software, and standard predictions, diagnostics, residuals and goodness of fit measures can be used.

Interpreting the effects of PC coordinates in a statistical model does not differ from common practice when using standard principal components, once coordinates themselves have been interpreted. The main difference is that PC coordinates in CoDA always imply trade-offs between eating more of some food group(s) and less of other(s), which does not need to be the case in standard principal component analysis.

The effects of balance coordinates or principal balances in a statistical model refer to the impact on the dependent variable when increasing all parts with positive exponents by a common factor and decreasing all parts with negative exponents by another common factor. For instance, a positive effect of the first balance coordinate y_1 on a health outcome would be interpreted as follows. Increasing vegetable, fruit, legume, seafood, cereal, and monounsaturated fat intake all by the same proportion while decreasing meat, dairy, and saturated fat intake all by the same proportion is related to a better health outcome.

Thus, in the context of dietary research, CoDA puts emphasis on the fact that any dietary pattern constitutes a trade-off between eating more of some foods and less of others. The relative importance of food groups, nutrients, or a combination of both, lies at the core of the research interest. No pattern derived by CoDA will ever imply eating more of all food groups or less of all food groups. This nicely fits both the intuitive notion of dietary pattern and usual practice in dietary recommendation.

CoDA offers a diverse toolbox which enables researchers to benefit from the best features of data-driven and investigator-driven methods. The closest to being a data-driven approach are PC coordinates extracted from compositional PC analysis, which is carried out in the same way as standard PC, once data have been appropriately transformed. On the other hand, the closest to being an investigator-driven approach are balance coordinates, in other words, log-ratios of geometric means of dietary components which the investigator wishes to compare or relate. The first balance coordinate is an attractive substitute for classical indexes of adherence to the MD such as Trichopoulou's, which are discrete.[4] The balance coordinate has (i) continuous unbounded distribution that may better lend itself to classic statistical models with, for instance, normally distributed variables, and (ii) computation not reliant on sample-derived medians, tertiles and the like. However, adherence to the MD admittedly does not explain a large proportion of variance in dietary compositional patterns in our control sample. In the opposite extreme, including all D-1 investigator-driven balance coordinates as predictors of any health outcome would take into account all variance in dietary patterns at the expense of parsimony.

The recently developed principal balances share some features with data-driven methods and others with investigator-driven methods. Like the investigator-driven method, they can be understood as log-ratios of geometric means of dietary components, with the added attractive property that most of the variance concentrates on a few principal balances, which enables parsimonious models when used as explanatory variables. Like the data-driven method, they can be understood as an equivalent to PC coordinates constrained to have equal loadings, which are easier to interpret and to replicate in comparative research. Rather than suggesting that one approach is superior under all circumstances, each method is designed to answer questions in a different way. When predicting a health outcome, data-driven analysis focuses on the variation in intakes, whereas investigator-driven analysis focuses on predefined dietary guidelines. Each approach has unique strengths and limitations, and their relative merits can ultimately depend on how well they predict each particular health outcome. In some cases, the patterns with the highest explanatory power on a health outcome may not be those with the highest variance or those based on previous theoretical knowledge. In such cases, using all D-1 balance coordinates may be more appropriate than using just the first few PC coordinates, principal balances or one or several investigator-driven indexes.[11,12,21,24]

As regards limitations of the proposed approaches, it must also be taken into account that, in spite of some attempts,[29,58] CoDA is not fully developed for sparse data tables, in other words, those with large proportions of zeros,[46] which would be the case when subdividing food groups in great detail (e.g. separated weekly intake of beef, pork, rabbit, lamb, horse, poultry and other meats). The case of structural zeros is also currently underdeveloped in CoDA[59] and it can be problematic to treat nondrinkers, vegetarians, vegans or even subjects of certain religions. Two limitations concern the present study rather than the methods themselves. First, our analysis was based on classical CoDA, but robust alternatives are available.[24,39] Second, for simplicity purposes, the illustration has been restricted to the subset of food groups and nutrients which are relevant to MD, but the whole set of groups available from the FFQ could have been used.[60]

The goal of dietary pattern analysis is to examine the multiple dimensions of the diet simultaneously relative to a given outcome. In this respect, CoDA provides an interesting alternative perspective. The proposed approaches

seem to hold promise for investigating the relationships between dietary patterns and diseases, given the compositional nature of research questions about diet.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was partially funded by the ''Accion Transversal del Cancer'', approved by the Spanish Ministry Council on 11 October 2007; by the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP); by the Instituto de Salud Carlos III-FEDER (grant nos PI08/1770, PI08/0533, PI08/1359, PS09/00773-Cantabria, PS09/01286-León, PS09/01903-Valencia, PS09/02078-Huelva, PS09/01662-Granada, PI11/01403, PI11/01810, PI11/02213, PI12/00488, PI14/01219, PI14/0613, PI15/00069, PI15/00914, PI15/01032, PI17/01280, PI09/0914, IJCI-2014-20900); by the Spanish Ministry of Health (grant no. CB06/02/1002); by the Spanish Ministry of Economy and Competitiveness (grant no. MTM2015-65016-C2-1-R); by the Catalan Government-Agency for Management of University and Research Grants (AGAUR) (grant nos 2014SGR551, 2017SGR656, 2017SGR733, 2017SGR723, 2017SGR1085); by the University of Girona (grant no. MPCUdG2016/069, GDRCompetUdG2017/19); by the Fundación Marqués de Valdecilla (grant no. API 10/09); by the Junta de Castilla y León (grant no. LE22A10-2); by the Consejería de Salud of the Junta de Andalucía (grant nos PI-0571-2009, PI-0306-2011, salud201200057018tra); by the Conselleria de Sanitat of the Generalitat Valenciana (grant no. AP_061/10); by the Regional Government of the Basque Country; by the Consejería de Sanidad de la Región de Murcia; by the European Commission (FOOD-CT-2006-036224-HIWATE); by the Spanish Association Against Cancer (AECC) Scientific Foundation; by the Fundación Caja de Ahorros de Asturias; and by the University of Oviedo. ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya.

ORCID iD

M Saez http://orcid.org/0000-0003-1882-0157

References

1. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002; **13**: 3–9.
2. Jacques PF and Tucker KL. Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr* 2001; **73**: 1–2.
3. Reedy J, Wirfalt E, Flood A, et al. Comparing 3 dietary pattern methods–cluster analysis, factor analysis, and index analysis – with colorectal cancer risk: the NIH-AARP Diet and Health Study. *Am J Epidemiol* 2010; **171**: 479–487.
4. Mila-Villarroel R, Bach-Faig A, Puig J, et al. Comparison and evaluation of the reliability of indexes of adherence to the Mediterranean diet. *Public Health Nutr* 2011; **14**: 2338–2345.
5. Castelló A, Buijsse B, Martin M, et al. Evaluating the applicability of data-driven dietary patterns to independent samples with a focus on measurement tools for pattern similarity. *J Acad Nutr Diet* 2016; **116**: 1914–1924.
6. Castelló A, Lope V, Vioque J, et al. Reproducibility of data-driven dietary patterns in two groups of adult Spanish women from different studies. *Br J Nutr* 2016; **116**: 734–742.
7. Pawlowsky-Glahn V, Egozcue JJ and Tolosana-Delgado R. *Modeling and analysis of compositional data*. Chichester: Wiley, 2015, pp.1–247.
8. Pedišić Ž, Dumuid D and Olds TS. Integrating sleep, sedentary behaviour, and physical activity research in the emerging field of time-use epidemiology: definitions, concepts, statistical methods. *Kinesiol Int J Fundam Appl Kinesiol* 2017; **49**: 10–11.
9. Chastin SFM, Palarea-Albaladejo J, Dontje ML, et al. Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: a novel compositional data analysis approach. *PLoS One* 2015; **13**: e0139984.
10. Carson V, Tremblay MS, Chaput J-P, et al. Associations between sleep duration, sedentary time, physical activity, and health indicators among Canadian children and youth using compositional analyses. *Appl Physiol Nutr Metab* 2016; **41**: 294–302.
11. Dumuid D, Pedišić Ž, Stanford TE, et al. The compositional isotemporal substitution model: a method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour. *Stat Methods Med Res*. Epub ahead of print 1 January 2017. doi: 10.1177/0962280217737805
12. Dumuid D, Stanford TE, Martin-Fernández J-A, et al. Compositional data analysis for physical activity, sedentary time and sleep research. *Stat Methods Med Res*. Epub ahead of print 1 January 2017. doi: 10.1177/0962280217710835

13. Dumuid D, Olds T, Lewis LK, et al. Health-related quality of life and lifestyle behavior clusters in school-aged children from 12 countries. *J Pediatr* 2017; **183**: 178–183.
14. Dumuid D, Maher C, Lewis LK, et al. Human development index, children's health-related quality of life and movement behaviors: a compositional data analysis. *Qual Life Res* 2018; **27**: 1473–1482.
15. Dumuid D, Lewis LK, Olds TS, et al. Relationships between older adults' use of time and cardio-respiratory fitness, obesity and cardio-metabolic risk: a compositional isotemporal substitution analysis. *Maturitas* 2018; **110**: 104–110.
16. Hunt T, Williams MT, Olds TS, et al. Patterns of time use across the chronic obstructive pulmonary disease severity spectrum. *Int J Environ Res Public Health* 2018; **15**: 533.
17. Foley L, Dumuid D, Atkin AJ, et al. Patterns of health behaviour associated with active travel: a compositional data analysis. *Int J Behav Nutr Phys Act* 2018; **15**: 26.
18. Talarico R and Janssen I. Compositional associations of time spent in sleep, sedentary behavior and physical activity with obesity measures in children. *Int J Obes* (Lond). Epub ahead of print 5 March 2018. doi: 10.1038/s41366-018-0053-x
19. Gupta N, Mathiassen SE, Mateu-Figueras G, et al. A comparison of standard and compositional data analysis in studies addressing group differences in sedentary behavior and physical activity. *Int J Behav Nutr Phys Act* 2018; **15**: 53.
20. Decarli A and Ferraroni M. Compositional data analysis and diversity indices: different approaches to define the role of nutrients in the study of diet-cancer relationship. In: *Proceedings from the first international conference on dietary assessment methods*, St. Paul, MN, 20–23 September 1992.
21. Leite MLC. Applying compositional data methodology to nutritional epidemiology. *Stat Methods Med Res* 2016; **25**: 3057–3065.
22. Leite MLC and Prinelli F. A compositional data perspective on studying the associations between macronutrient balances and diseases. *Eur J Clin Nutr* 2017; **71**: 1365–1369.
23. Trinh HT, Morais J, Thomas-Agnan C, et al. Relations between socio-economic factors and nutritional diet in Vietnam from 2004 to 2014: new insights using compositional data analysis. *Stat Methods Med Res* 2018 (in press). Epub ahead of print 1 January 2018. doi: 10.1177/0962280218770223
24. Mert MC, Filzmoser P, Endel G, et al. Compositional data analysis in epidemiology. *Stat Meth Med Res* 2018; **27**: 1878–1891.
25. Quinn TP, Erb I, Richardson MF, et al. Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* 2018 (in press). doi: 10.1093/bioinformatics/bty175
26. Friedman J and Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 2012; **8**: e1002687.
27. Pinto JR, Egozcue JJ, Pawlowsky-Glahn V, et al. Balances: a new perspective for microbiome analysis. *bioRxiv*. Epub ahead of print 2017. doi: 10.1101/219386
28. Tsilimigras MCB and Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann Epidemiol* 2016; **26**: 330–335.
29. Kurtz ZD, Müller CL, Miraldi ER, et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol* 2015; **11**: e1004226.
30. Aitchison JA. The statistical analysis of compositional data. *J R Stat Soc Ser B* 1982; **44**: 139–177.
31. Aitchison JA. *The statistical analysis of compositional data. Monographs on statistics and applied probability*. London: Chapman and Hall, 1986, p.416.
32. Buccianti A, Mateu-Figueras G and Pawlowsky-Glahn V. *Compositional data analysis in the geosciences: from theory to practice*. London: Geological Society, 2006, p.212.
33. Barceló-Vidal C and Martín-Fernández JA. The mathematics of compositional analysis. *Austrian J Stat* 2016; **45**: 57–71.
34. Azevedo Rodrigues L, Daunis-i-Estadella J, Mateu-Figueras G, et al. Flying in compositional morphospaces: Evolution of limb proportions in flying Vertebrates. In: Pawlowsky-Glahn A and Buccianti V (eds) *Compositional data analysis: theory and applications*. Chichester: John Wiley & Sons, Ltd, 2011, pp.235–254.
35. Trichopoulou A, Costacou T, Bamia C, et al. Adherence to a Mediterranean diet and survival in a Greek population. *N Engl J Med* 2003; **348**: 2599–2608.
36. van den Boogaart KG and Tolosana-Delgado R. *Analyzing compositional data with R*. Heidelberg: Springer, 2013, pp.1–258.
37. Palarea-Albaladejo J and Martín-Fernández JA. ZCompositions – R package for multivariate imputation of left-censored data under a compositional approach. *Chemom Intell Lab Syst* 2015; **143**: 85–96.
38. Quinn TP, Richardson MF, Lovell D, et al. propr: An R-package for identifying proportionally abundant features using compositional data analysis. *Scientific Rep* 2017; **7**: 16252.
39. Templ M, Hron K and Filzmoser P. robCompositions: an R-package for robust statistical analysis of compositional data. In: Pawlowsky-Glahn A and Buccianti V (eds) *Compositional data analysis: theory and applications*. Chichester: John Wiley & Sons, Ltd, 2011, pp.341–55.
40. Thió-Henestrosa S and Martín-Fernández JA. Dealing with compositional data: the freeware CoDaPack. *Math Geol* 2005; **37**: 773–793.

41. Pawlowsky-Glahn A and Buccianti V. *Compositional data analysis: theory and applications*. Chichester: John Wiley & Sons, Ltd, 2011, pp.1–378.
42. Aitchison J. Principal component analysis of compositional data. *Biometrika* 1983; **70**: 57–65.
43. Aitchison J. Simplicial inference. In: Marlos AGV and Richards DSP (eds) *Algebraic methods in statistics and probability: AMS special session on algebraic methods in statistics*. Providence: American Mathematical Society, 2001, pp.1–22*Contemporary mathematics series*.
44. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, et al. Isometric logratio transformations for compositional data analysis. *Math Geol* 2003; **35**: 279–300.
45. Lovell D, Pawlowsky-Glahn V, Egozcue JJ, et al. Proportionality: A valid alternative to correlation for relative data. *PLoS Comput Biol* 2015; **11**: e1004075.
46. Martín-Ferńandez JA, Palarea-Albaladejo J and Olea RA. Dealing with zeros. In: Pawlowsky-Glahn A and Buccianti V (eds) *Compositional data analysis: theory and applications*. Chichester: John Wiley & Sons, Ltd, 2011, pp.43–58.
47. Palarea-Albaladejo J and Martín-Fernández JA. A modified EM alr-algorithm for replacing rounded zeros in compositional data sets. *Comput Geosci* 2008; **34**: 902–917.
48. Bruno F, Greco F and Ventrucci M. Spatio-temporal regression on compositional covariates: modeling vegetation in a gypsum outcrop. *Environ Ecol Stat* 2015; **22**: 445–463.
49. Mateu-Figueras G, Pawlowsky-Glahn V and Díaz-Barrero JL. The principle of working on coordinates. In: Pawlowsky-Glahn A and Buccianti V (eds) *Compositional data analysis: theory and applications*. Chichester: John Wiley & Sons, Ltd, 2011, pp.29–42.
50. Egozcue JJ and Pawlowsky-Glahn V. Groups of parts and their balances in compositional data analysis. *Math Geol* 2005; **37**: 795–828.
51. Pawlowsky-Glahn V and Egozcue JJ. Exploring compositional data with the CoDa-Dendrogram. *Austrian J Stat* 2016; **40**: 103–113.
52. Martín-Fernández JA, Pawlowsky-Glahn V, Egozcue JJ, et al. Advances in principal balances for compositional data. *Math Geosci* 2018; **50**: 273–298.
53. Pawlowsky-Glahn V, Egozcue JJ and Tolosana-Delgado R. Principal balances. In: *The 4th international workshop on compositional data analysis CoDaWork2011*, Sant Feliu de Guíxols, Spain, 9–23 May 2011. Girona: University of Girona, pp.1–10.
54. Castano-Vinyals G, Aragones N, Perez-Gomez B, et al. Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. *Gac Sanit* 2015; **29**: 308–315.
55. Filzmoser P and Hron K. Outlier detection for compositional data using robust methods. *Math Geosci* 2008; **40**: 233–248.
56. Tolosana-Delgado R and van den Boogaart KG. Linear models with compositions in R. In: Pawlowsky-Glahn A and Buccianti V (eds) *Compositional data analysis: theory and applications*. Chichester: John Wiley & Sons, Ltd, 2011, pp.356–371.
57. Coenders G, Martín-Fernández JA and Ferrer-Rosell B. When relative and absolute information matter. Compositional predictor with a total in generalized linear models. *Stat Model* 2017; **17**: 494–512.
58. Mert MC, Filzmoser P and Hron K. Sparse principal balances. *Stat Model An Int J* 2015; **15**: 159–174.
59. Roel V, Antonio K and Claeskens G. Unraveling the predictive power of telematics data in car insurance pricing, http://dx.doi.org/10.2139/ssrn.2872112 (accessed 7 November 2017).
60. Castello A, Boldo E, Perez-Gomez B, et al. Adherence to the Western, Prudent and Mediterranean dietary patterns and breast cancer risk: MCC-Spain study. *Maturitas* 2017; **103**: 8–15.